THE PROCEEDINGS OF
THE PHYSICAL SOCIETY

VOL. 61, PART 2 1 August 1948 No. 344

Shock Waves in Air at Very High Pressures *

BY D. R. DAVIES

Communicated by W. G. Penney ; MS. received 5 *September* 1947, *and in amended form* 18 *February* 1948 ; *read* 9 *January* 1948

ABSTRACT. The equations of the shock wave in air at N.T.P. are solved numerically for pressures on the high pressure side up to 1000 atmospheres. The solution enables all the physical entities, such as temperature, wave velocity, particle velocity, air density, etc., on the shock wave front to be expressed numerically as a function of the wave pressure. To solve these equations it is necessary to know the internal energy (E) and volume (v) of one gramme of air over certain regions of the two-dimensional range $1<p<1000$ atmospheres and $273<T<16\,000°$ K. Calculations are carried out to assess the numbers of the various types of molecules, atoms and ions present at any p and T. The only E values needed at high p are those for which T is also high, and the simple gas laws may therefore be assumed for the volume determinations. The resultant energy is obtained by summing the internal energies of the different groups of atoms and molecules present. Statistical mechanics furnishes equations whose solution fixes the composition. This set of equations is difficult to solve to any degree of accuracy if oxides of nitrogen are taken into account ; of these only nitric oxide, NO, ever exceeds a concentration of 1% by weight, and its maximum concentration is less than 5%. As a first approximation, the shock wave equations are worked out on the assumption that the species present are N_2, O_2, N and O. These calculations make full allowance for all quantum states apart from ionic states. They are refined later by taking into consideration the presence of NO and argon (1·3% by weight) and ionization possibilities.

§1. INTRODUCTION

DURING the early part of the recent war the problem of working out theoretically the distribution of pressure and velocity *behind* a spherical shock wave in water was resolved by solving the hydrodynamical equations for the motion of water behind the waves by numerical methods. The similar problem of propagation of shock waves in *air*, however, proved to be much more difficult, as the equation of state of air at high temperatures presented considerable difficulties. The first step in the solution consisted of working out the relationship between shock wave velocity, mass velocity, temperature, pressure, etc., on the shock wave itself, assuming a sudden discontinuity at the wave front. It is this problem which is described in the paper.

A system of equations was derived by Lewis and von Elbe (1938) relating the various physical entities connected with a shock wave front. The notation used in the paper for these physical entities is as follows : shock wave velocity U, mass velocity u, temperature T, volume of unit mass of air v, internal energy per unit mass of air E, and the air pressure p. These shock wave equations may be solved if we have a complete knowledge of the *internal energy of air* in the neighbourhood of the shock wave conditions. They completely determine

* This paper forms part of a thesis presented to the University of London for the Ph.D. degree.

ail the variables U, u, T, p, v, etc., *if one of the set is known*, in terms of the known conditions ahead of the wave front. It is assumed in this paper that the air on the low pressure side is at N.T.P. The independent variable chosen is p, the pressure on the high pressure side. Provided that only differences of pressure amounting to not more than a few atmospheres on the two sides of the wave are considered, simple calculations lead to reliable values of U, u, and T, etc., in terms of p, as worked out by Lewis and von Elbe.

Solution becomes difficult for p values above five atmospheres due to the variation of the specific heats of air from the values obtained at ordinary temperatures and pressures, as very high temperatures are generated on the wave front at high pressures. For pressures up to about thirty atmospheres the only complication to the simple theory arises out of the necessity to allow for the vibrational energy of air molecules. This paper is concerned with the pressure range from thirty to one thousand atmospheres, where, due to several complications, an elaborate calculation becomes necessary. Dissociation of the air molecules must be taken into account: molecules and atoms appear on electronic energy levels other than their ground states, and allowances must be made, too, for the changing composition of the air as the pressure and temperature conditions change. At pressures greater than one thousand atmospheres ionization complicates matters further, although dissociation has ceased to be troublesome. These calculations are not treated in this paper.

It becomes necessary as a first step to evaluate the internal energy E and the volume v of one gramme of air over most of the range $1<p<1000$ atmospheres and $273<T<16\,000°$ K. At this first stage in the solution, standard methods of statistical mechanics furnish equations whose solution fixes the composition of the air. If great accuracy is required, these equations provide numerical difficulties. In addition to the formation of nitrogen and oxygen atoms, argon and oxides of nitrogen have to be dealt with. However, only nitric oxide, NO, and argon, A, ever appear in concentrations greater than 1% by weight. Once the composition has been fixed, the volume follows immediately from the law $p=nkt$; and the internal energy is the sum of the internal energies of the different groups of atoms and molecules present.

As a first approximation only, nitrogen and oxygen molecules and atoms are considered, and E and v are evaluated over the range specified above. These functions enable the shock wave equations to be solved and thus specify approximately the neighbourhood within which the wave conditions lie. These calculations are later refined by taking into account (*a*) formation of NO, (*b*) formation of argon and (*c*) a more recent evaluation by A. G. Gaydon and W. G. Penney (1945) of the dissociation energy of nitrogen. This later work is greatly simplified, as the results of the first approximation considerably narrow the two-dimensional range of T and p, within which the numerical work has to be carried out.

§ 2. INTERNAL ENERGY PER GRAMME FOR THE MAIN CONSTITUENTS OF AIR UP TO 16 000° K.

Air at normal temperatures and pressures consists primarily of oxygen and nitrogen molecules with a small proportion of inert gases, mainly argon. The percentages by weight are: nitrogen 75·5, oxygen 23·2, argon 1·3. At a temperature of about 2000° K. the oxygen commences to dissociate, and near 3000° K. the nitrogen constituent behaves similarly. At the 4000° K. stage a considerable

proportion of the nitrogen atoms combine with oxygen atoms to form oxides of nitrogen. Of these only nitric oxide ever exceeds the figure of 1% by weight. The components considered in this calculation are nitrogen and oxygen atoms and molecules, nitric oxide and argon. The proportion by weight of these gases, in one gramme of air, over the temperature range 4000–16000° K. varies considerably with temperature and pressure conditions. It becomes necessary, therefore, at the outset, to evaluate the energy per gramme of these components over the temperature range and later to assess the degrees of dissociation at each temperature.

Statistical thermodynamic theory leads to simple formulae for the translational, rotational, vibrational and electronic energies of the various components. These are: (i) the internal energy per molecule, or atom, associated with the three-dimensional degrees of freedom

$$E_t = 3\boldsymbol{k}T/2, \qquad \ldots\ldots(1)$$

where $\boldsymbol{k}$ = Boltzmann's constant; (ii) the internal energy per molecule of the two rotational degrees of freedom of a diatomic molecule,

$$E_r = \boldsymbol{k}T; \qquad \ldots\ldots(2)$$

(iii) the internal energy of vibration of a molecule, per molecule,

$$E_v = \boldsymbol{h}\nu/[\exp(\boldsymbol{h}\nu/\boldsymbol{k}T) - 1], \qquad \ldots\ldots(3)$$

where ν is the fundamental vibration frequency and $\boldsymbol{h}$ is Planck's constant; (iv) the internal energy per molecule or atom, associated with the electronic motion

$$E_e = \{\Sigma_i g_i W_i \exp(-W_i/\boldsymbol{k}T)\}/\{\Sigma_i g_i \exp(-W_i/\boldsymbol{k}T)\}, \qquad \ldots\ldots(4)$$

where g_i is the statistical weight and W_i the energy of the electronic level i. The ground state of an atom is taken to be $D/2$, where D is the energy of dissociation of the ground state of the molecule formed from two of the atoms.

These formulae, and later ones covering the equilibrium constants which specify the degrees of dissociation, are based on certain necessary approximations. It is assumed that complete factorization of the partition functions, § 3, upon which the energy functions and dissociation formulae are based, is exact, hence that the total energy is the sum of the separate energies. This is not strictly true. The spacing of the vibrational levels is not constant, and the moment of inertia of the molecule changes with increasing vibration and from one electronic level to another. However, the errors introduced by these approximations are not large, especially in view of the order of accuracy required in the solution of the shock wave equations.

Table 1 gives the spectroscopic data used in the energy and dissociation formulae.

Table 1

	D (ev.)	ν (cm^{-1})	B_e (cm^{-1})
O_2	5·082	1580	1·440
N_2	9·764	2360	1·998
NO	6·490	1900	1·684

The relation between B_e and I, the moment of inertia, is $B_e = \boldsymbol{h}/(8\pi^2 I\boldsymbol{c})$, where c is the velocity of light (3×10^{10} cm/sec.).

Other spectroscopic data that we require may be summarized. The ground state of O is $^3\Sigma_1$; there is another state $^1\Delta$, at 0·96 ev., and a second excited state $^1\Sigma$, at 1·62 ev. The ground state of O is ^3P; there are excited ^1D at 1·97 ev. and ^1S at 4·5 ev. The ground state of N_2 is $^1\Sigma$ and the ground state of N is ^4S. The excited states of N_2 and N are too high to be considered as they have very small Boltzmann factors. These $^-\Sigma$ and ^1S states of O_2 and O have little effect, but they have been included. The ground state of NO is $^2\Pi$ and the first excited state $^2\Sigma$. High excited levels have no significant influence on the energy functions.

The statistical weights of the excited states may be written down once the electronic levels are known. For example, statistical weight of the ground state of O is 9, because the level has spin degeneracy 3 (triplet nature) and orbital degeneracy 3 (P states). The multiplet structure has been neglected because the overall separation of the ^3P multiplet is somewhat less than kT at temperatures producing appreciable dissociation of the oxygen molecules.

Statistical weights used for the ground state, first and second excited levels respectively are: for O_2—3, 5, 1, for O—9, 5, 1 and for NO—2 and 2 for the ground state and first excited level.

Table 2. Internal energy of air constituents

Substance	T (° K.)	$E_t+E_r+E_v$ (cal/gm.mol.)	Electronic energy (cal/gm.mol.)	Total internal energy (cal/gm.)
Oxygen molecules	2,000	12,060	142	381
	6,000	39,600	4,975	1,393
	10,000	67,330	8,800	2,379
Oxygen atoms	4,000	11,910	86	750
	8,000	23,830	1,454	1,580
	12,000	35,750	3,620	2,460
	16,000	47,660	5,722	3,337
Nitrogen molecules	2,000	11,450	0	409
	6,000	38,710	0	1,382
	10,000	66,300	0	2,368
	14,000	94,124	0	3,362
Nitrogen atoms	2,000	5,960	0	426
	6,000	17,870	0	1,277
	10,000	29,790	0	2,128
	14,000	41,707	0	2,978
Nitric oxide	2,000	11,794	0	393
	6,000	39,150	4	1,305
	10,000	66,930	227	2,238
	14,000	94,696	1,366	3,202
Argon	2,000	5,957	0	149
	6,000	17,874	0	448
	10,000	29,790	0	746
	14,000	41,707	0	1,193

Note: Intermediate values for this and ensuing tables may be obtained on application to the author.

The formulae used for the electronic energies in cal/gm. atom or molecule: for oxygen molecules,

$$E_{el}=23{\cdot}07\times10^3(4{\cdot}8\,e^{-x_1}+1{\cdot}62\,e^{-x_2})/(3+5\,e^{-x_1}+e^{-x_2}); \quad \ldots\ldots(5)$$

for oxygen atoms,

$$E_{el} = 23{\cdot}07 \times 10^3 (9{\cdot}835\, e^{-y_1} + 4{\cdot}5\, e^{-y_2})/(9 + 5e^{-y_1} + e^{-y_2}); \quad \ldots\ldots(6)$$

for nitric oxide molecules,

$$E_{el} = 23{\cdot}07 \times 10^3 (5{\cdot}453\, e^{-z_1})/(1 + e^{-z_1}), \quad \ldots\ldots(7)$$

where $x_1 = P_1/\boldsymbol{k}T$, $x_2 = P_2/\boldsymbol{k}T$, $y_1 = Q_1/\boldsymbol{k}T$, $y_2 = Q_2/\boldsymbol{k}T$, $z_1 = R_1/\boldsymbol{k}T$, and P_1, Q_1, R_1 are the energies of the first excited levels and P_2, Q_2 the energies of the second excited levels, all in appropriate units. The relationship 23·07 cal/gm. atom = 1 ev. is used.

In table 2 the energies per gramme of each constituent of air are given as a function of the temperature.

§3. ESTIMATION OF THE EQUILIBRIUM CONSTANTS OF NITROGEN, OXYGEN AND NITRIC OXIDE UP TO 16 000° K.

It is possible with the aid of the general laws of thermodynamics to predict the form of the functional dependence of the degree of dissociation of a gas on the temperature and pressure conditions. These laws, however, leave unspecified certain constants. A knowledge of the energy levels which the particles of the gas can occupy, as furnished by standard spectroscopic methods, enables these constants to be evaluated; the degree of dissociation can thus be computed numerically from thermodynamics in conjunction with spectroscopic data.

The chemical reactions to be considered are:—

$$O_2 \rightleftharpoons O + O; \quad N_2 \rightleftharpoons N + N; \quad N + O \rightleftharpoons NO. \quad \ldots\ldots(8)$$

Many more, whose effect is small and have been neglected, may occur, such as

$$N_2 + O \rightleftharpoons N_2O; \quad N + O_2 \rightleftharpoons NO_2; \quad O \rightleftharpoons O^+ + e.$$

Consider the simplest type of dissociation equilibrium, $AB \rightleftharpoons A + B$, in which gaseous compound AB dissociates into the monatomic vapours A and B. Then the equilibrium constant K is defined by $K = p_A \cdot p_B/p_{AB}$, where p_A and p_B are the partial pressures of the monatomic vapours and p_{AB} the partial pressure of the original compound.

Using this definition in the nitrogen reaction $N_2 \rightleftharpoons 2N$, it is easily shown that

$$a_1 = \sqrt{(K_1/(K_1 + 4p_1))}, \quad \ldots\ldots(9)$$

or $\ln K_1 = \ln[4a_1^2 p_1/(1 - a_1^2)]$, where a_1 is the ratio = (weight of nitrogen atoms in a given volume)/(weight of nitrogen atoms plus molecules in the same volume), p_1 is the partial pressure of nitrogen ($N + N_2$), and K_1 the equilibrium constant of the reaction at temperature T.

A similar equation holds for oxygen (a_2). Hence, once the K value is fixed at a given temperature, the dissociation, given by a_1 etc., can be evaluated.

Calculation of equilibrium constant as a function of temperature is simple once the partition functions for the components have been obtained.

The partition function P_1 for O and N may both be expressed as a product of three linear translational partition functions and one electronic partition function $P_1 = T^3(m)P(A)$. A represents the atoms O or N, m the mass of the atom.

Similarly the P_2 functions for nitrogen and oxygen molecules may be written $P_2 = T^3(2m) \cdot R^2(I) \cdot V(\nu) \cdot P(M)/2$, where T represents translation, R represents

end-over-end rotations, V represents vibration, and P(M) represents the electronic part of the partition function of a molecule M. The factor $\frac{1}{2}$ enters for N_2 and O_2, because only one half of the rotational phase space is available to these homonuclear molecules N_2 and O_2. It does not enter for NO.

These standard expressions are used:

$$T(m)=v^3\sqrt{(2\pi m\boldsymbol{k}T)};\ R(I)=2\pi\sqrt{(2\pi I\boldsymbol{k}T)};$$
$$P(\mathrm{O})=9+5\exp(-Q_1/\boldsymbol{k}T)+\exp(-Q_2/\boldsymbol{k}T)+\ldots;$$
$$P(\mathrm{O_2})=3+5\exp(-P_1/\boldsymbol{k}T)+\exp(-P_2/\boldsymbol{k}T)+\ldots;$$
$$P(\mathrm{NO})=2+2\exp(-R_1/\boldsymbol{k}T)+\ldots;$$
$$P(\mathrm{N})=4;\ P(\mathrm{N_2})=1;\ V(\nu)=1/[1-\exp(-\boldsymbol{h}\nu/\boldsymbol{k}T)].$$

For the reaction $AB\rightleftharpoons A+B$, the equation obtained for $\ln K$ is

$$\ln K=-D/\boldsymbol{k}T+3/2\ln[2\pi M_A M_B\boldsymbol{k}T/\boldsymbol{h}^2M_{AB}]+\ln\boldsymbol{h}B_e+\ln\{1-\exp(-\boldsymbol{h}\nu/\boldsymbol{k}T)\}$$
$$+\ln g_0^A g_0^B/g_0^{AB}+\ln\frac{[1+(g_1^A/g_0^A)\exp(-X/\boldsymbol{k}T)][1+(g_1^B/g_0^B)\exp(-Y/\boldsymbol{k}T)]}{[1+(g_1^{AB}/g_0^{AB})\exp(-Z/\boldsymbol{k}T)]},\qquad\ldots\ldots(10)$$

where D = energy of dissociation from the lowest vibrational level; M_A, M_B = mass of atom A, B; $M_{AB}=(M_A+M_B)$ = mass of molecule AB; g_0, g_1 = statistical weight of atoms or molecules in the ground state and first level (the superscripts are self-explanatory); X, Y, Z = energy increase from ground state to the first level for atoms A, B, molecules AB respectively.

The second and higher electronic levels have no effect. On the basis of wave mechanics there must for N_2 and O_2 be added, to the right hand of the above equation, a term $\ln 2$ to allow for the symmetry of molecules AA.

The expression used for the oxygen reaction is

$$\ln K=-D_0/\boldsymbol{k}T+3/2\ln\pi m_0\boldsymbol{k}T/\boldsymbol{h}^2+\ln\boldsymbol{h}B_e+\ln[1-\exp(\boldsymbol{h}\nu_0/\boldsymbol{k}T)]+\ln 54$$
$$+2\ln[1+0{\cdot}556\exp(-Q_1/\boldsymbol{k}T)]-\ln[1+1{\cdot}667\exp(-P_1/\boldsymbol{k}T)],\qquad\ldots\ldots(11)$$

(m_0 = mass of oxygen atom). Similarly for nitrogen

$$\ln K=-D_N/\boldsymbol{k}T+3/2\ln\pi m_N\boldsymbol{k}T/\boldsymbol{h}^2+\ln\boldsymbol{h}B_e+\ln[1-\exp(-\boldsymbol{h}\nu_N/\boldsymbol{k}T)]+\ln 32\qquad\ldots\ldots(12)$$

(m_N = mass of nitrogen atoms), and for nitric oxide

$$\ln K=-D_{NO}/\boldsymbol{k}T+3/2\ln 2\pi m_N m_O\boldsymbol{k}T/\boldsymbol{h}^2m_{NO}+\ln\boldsymbol{h}B_e+\ln[1-\exp(-\boldsymbol{h}\nu_{NO}/\boldsymbol{k}T)]$$
$$+\ln 18+\ln[1+0{\cdot}556\exp(-Q_1/\boldsymbol{k}T)]-\ln[1+\exp\{-R_1/\boldsymbol{k}T\}].\qquad\ldots\ldots(13)$$

The B_e in this formula is $\boldsymbol{c}$ times that given in table 1, since frequencies are usually expressed in wave numbers as in table 1.

Table 3. Equilibrium constants of oxygen, nitrogen and nitric oxide as functions of temperature

T (° K.)	K (O) (dyne/cm²)	K (N) (dyne/cm²)	K (NO) (dyne/cm²)
2,000	0·595	0	0
4,000	$3{\cdot}509\times10^6$	3·390	$2{\cdot}754\times10^4$
8,000	$6{\cdot}740\times10^9$	$7{\cdot}959\times10^6$	$8{\cdot}715\times10^8$
12,000	$9{\cdot}488\times10^{10}$	$1{\cdot}158\times10^9$	$3{\cdot}623\times10^{10}$
16,000	$3{\cdot}929\times10^{11}$	$1{\cdot}458\times10^{10}$	$2{\cdot}620\times10^{11}$

§ 4. ESTIMATION OF INTERNAL ENERGY $E(p, T)$, AND VOLUME, $v(p, T)$, FOR ONE GRAMME OF AIR

In this section the effects of oxides of nitrogen and argon are neglected and a D value of 7·35 ev. is assumed for nitrogen. The general formulae used are

$$E=\sum_i n_i(p, T).E_i(T) \qquad \ldots\ldots(14)$$

and $\sum_i n_i(p, T)=n(p, T)$, where $n_i(p, T)$ is the number of any class of atoms or molecules i, $E_i(T)$ is the internal energy per particle of this type of particle, E the total internal energy per gramme of air, and

$$pv=n(p, T)\boldsymbol{k}T. \qquad \ldots\ldots(15)$$

Fortunately the only values of E required at high pressures are those for which the temperature is also high. At no point does v become less than about 1/60 of the volume at N.T.P. so that the use of equation (15) is justified.

By means of equations (9), (11) and (12), a_1 and a_2 may be evaluated over the dissociation region, and hence by (14) the energies per gramme of nitrogen and per gramme of oxygen as functions of p_1 or p_2 and T. It can be shown that

$$(1+a_1)/p_1=0{\cdot}269(1+a_2)/p_2, \qquad \ldots\ldots(16)$$

where p_1, p_2 are the partial pressures of nitrogen and oxygen respectively. Hence p_1 and p_2 may be evaluated at each temperature for the required value of total pressure $p(=p_1+p_2)$. Interpolating at these p_1 and p_2 values in the energy curves for oxygen and nitrogen, and assuming that one gramme of air always contains 24% by weight of oxygen and 76% by weight of nitrogen, we obtain the final curves for the internal energy per gramme of air for $273<T<16000°$ K. and total pressures equal to 15, 25, 35, 45, 65, 75, 150, 200, 400, 600, 800 and 1000 atmospheres; thus later interpolation between two p curves is easily performed.

The general character of these results is seen in figure 1. Once the p_1 and p_2 values have been determined for a given p, the volume of one gramme of air is given by

$$v=\frac{\frac{1}{2}\boldsymbol{k}T}{(p_1 m_N/(1+a_1))+(p_2 m_O/(1+a_2))}. \qquad \ldots\ldots(17)$$

Some of these values are indicated in figures 1 and 2.

It must be borne in mind, however, that these curves are intended to give an indication only of the actual trend of the E and v curves. They depend on a previous determination of D for nitrogen and neglect the formation of oxides of nitrogen and the presence of inert gases, etc.

§ 5. SOLUTION OF THE SHOCK WAVE EQUATIONS: FIRST APPROXIMATION

Consider a coordinate system moving with the wave front. The mechanism of transfer across the wave front is stationary, and consequently the mass, momentum and energy of a gas passing across the wave front must be conserved. According to Lewis and von Elbe (1938),

$$u_1/v_1=u_2/v_2;\ \ p_1+u_1^2/v_1=p_2+u_2^2/v_2;\ \ E_1+u_1^2/2+p_1v_1=E_2+u_2^2/2+p_2v_2.$$

From these it follows that

$$\Delta E=\tfrac{1}{2}(p_2+p_1)(v_1-v_2), \qquad \ldots\ldots(18)$$

$$U=u_1=v_1\sqrt{\{(p_2-p_1)/(v_1-v_2)\}}, \qquad \ldots\ldots(19)$$

$$u=u_1-u_2=\sqrt{\{(p_2-p_1)(v_1-v_2)\}}, \qquad \ldots\ldots(20)$$

where p_1, v_1 refer to the low-pressure side and p_2, v_2 to the high-pressure side.

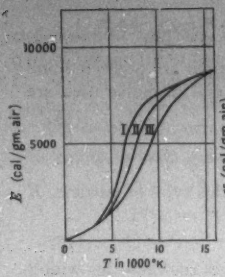


Figure 1. Internal energy of air at constant pressure for varying T, neglecting the presence of NO and A. $D_0 = 7{\cdot}35$ ev.; I, $p = 15$ atmos.; II, $p = 150$ atmos.; III, $p = 1000$ atmos.

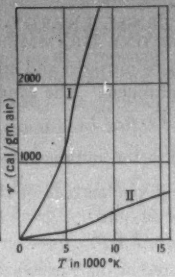


Figure 2. Volume of air at constant pressure for varying T, on assumptions of figure 1.

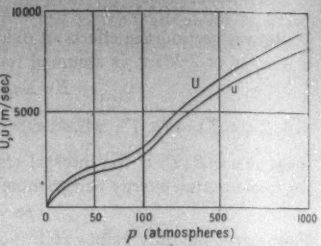


Figure 3. Shock wave velocity U and mass velocity u as a function of p. (Note: two scales for p.)

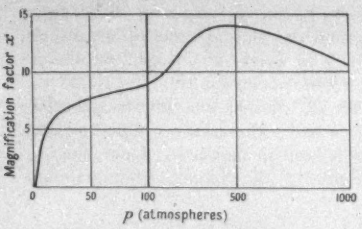


Figure 4. Magnification factor x. (Note: two scales for p.)

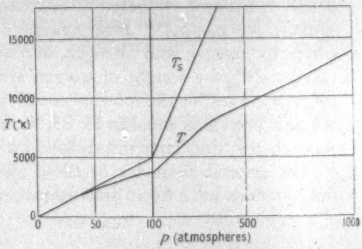


Figure 5. The temperature T behind the shock wave. The temperature T_s is also shown, where T_s is calculated on the assumption that $\gamma = 1{\cdot}4$ for all T and p. Two scales are used.

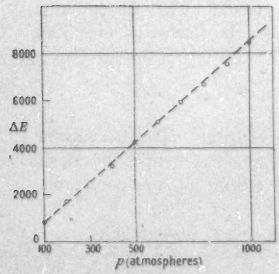


Figure 6. The change ΔE in the internal energy per gramme of air, due to the passage of a shock wave, as a function of the pressure on the wave. ΔE is in calories per gramme of air; p in atmospheres.

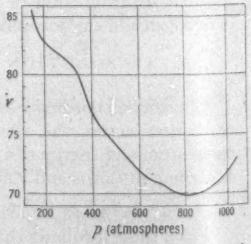


Figure 7. The volume (in cm³) of one gramme of air as a function of the shock wave pressure (in atmos.).

These equations may be solved numerically by a process of successive approximation, starting with p_2 and assessing a value for v_2. The volume and energy curves then give T_2 and E_2. If on one side the air is at N.T.P., E_1 is known, and hence ΔE. By comparison with the value of ΔE calculated from (18), adjustment may be made to v_2. Values of ΔE, T_2, U and u follow directly. These are given in table 4.

Table 4. Solution of the shock wave equations (neglecting oxides of nitrogen and argon; and assuming D for nitrogen to be 7·35 ev.)

p_2 (atmospheres)	30	100	500	1,000
U (metres/sec.)	1,670	3,000	6,560	9,350
u (metres/sec.)	1,370	2,590	5,960	8,310
T_2 (° K.)	1,480	3,600	7,950	15,000
E_2 (calories per gm. of air)	287	875	4,320	8,450
v_2 (cm^3 per gm. of air)	140	107	70	86

§6. SOLUTION OF THE SHOCK WAVE EQUATIONS: INCORPORATING EFFECT OF NO AND A

In this solution the value of D for nitrogen is 9·764.

It has been shown that the effect of allowing for the formation of NO is a small one, resulting at its maximum (4000–5000° K.) in an increase in the value of E_2 of about 12 calories and a decrease of p_2 of about 5 atmospheres. Allowance for argon has a similarly small effect. The increase in the nitrogen dissociation value has a much larger influence on the shock wave conditions. The first solution is unchanged up to about 100 atmospheres.

The dissociation equations are easily written down. Consider the species O_2, O, N_2, N, NO and argon. The last component does not change, so that its influence is easily predicted. If v is the volume of one gramme of air at temperature T, α the number of oxygen atoms, present as O, O_2 or in NO, in one gramme of air, β the number of nitrogen atoms, present as N, N_2 or in NO, in one gramme of air, x, y the fraction of oxygen, nitrogen, in the form of atoms, and z is the number of molecules of NO, the equations relating x, y, z, v and p are

$$x^2=[1-x-z]\,vK_1/2\alpha\boldsymbol{k}T, \qquad \ldots\ldots(21)$$

$$y^2=[1-y-z\alpha/\beta]\,vK_2/2\beta\boldsymbol{k}T, \qquad \ldots\ldots(22)$$

$$z=(xy/v)\beta\boldsymbol{k}T/K_3, \qquad \ldots\ldots(23)$$

$$p'=(\boldsymbol{k}T/2v)[\alpha(x+1)+y(\beta+1)], \qquad \ldots\ldots(24)$$

$$p=p'+p_A, \qquad \ldots\ldots(25)$$

where K_1, K_2, K_3 are the equilibrium constants at temperature T for oxygen, nitrogen and nitric oxide respectively, p_A = partial pressure of the argon and p' = partial pressure of remaining constituents.

From (21), (22) and (23),

$$x=\frac{(1-z)(1-\alpha z/\beta)K_2/2K_3-2\alpha K_3z^2/\beta K_1}{(1-\alpha z/\beta)K_2/2K_3+\alpha K_2z/\beta K_1}, \qquad \ldots\ldots(26)$$

and

$$y=\frac{(1-z)(1-\alpha z/\beta)K_2/2K_3-2\alpha K_3z^2/\beta K_1}{(1-z)K_2/2K_3+z}. \qquad \ldots\ldots(27)$$

Substitution in (24) gives a bi-quadratic for z:

$$Az^4+Bz^3+Cz^2+Dz+E=0, \qquad \ldots\ldots(28)$$

where

$$A=+\left(\frac{\alpha}{\beta}\right)^2\left[4\left(\frac{K_3}{K_1}\right)^2-2\left(\frac{K_2}{K_1}\right)+\frac{1}{4}\left(\frac{K_2}{K_3}\right)^2\right]$$
$$-\left(\frac{\alpha}{\beta}\right)^2\left(\frac{K_3}{2p'}\right)\left(\frac{K_2}{2K_3}-\frac{2K_3}{K_1}\right)\left(1-\frac{K_2}{K_3}+\frac{K_2}{K_1}\right);$$

$$B=+\left(\frac{\alpha}{\beta}\right)\left(1+\frac{\alpha}{\beta}\right)\left\{\frac{2K_2}{K_1}-\frac{1}{2}\left(\frac{K_2}{K_3}\right)^2\right\}-\left(\frac{\alpha}{\beta}\right)\left(\frac{K_3}{2p'}\right)\left[\left(\frac{\alpha}{\beta}+1\right)\left(\frac{K_2}{K_3}\right)\left(\frac{K_2}{2K_3}-\frac{2K_3}{K_1}\right)\right]$$
$$+\left(\frac{\alpha}{\beta}\right)\left(\frac{K_3}{2p'}\right)\left[\left(1-\frac{K_2}{2K_3}\right)\left(\frac{\alpha K_2}{\beta K_1}-\frac{\alpha K_2}{\beta K_3}-\frac{K_2}{2K_3}\right)\right.$$
$$\left.+\left(1-\frac{K_2}{K_3}-\frac{\alpha K_2}{2\beta K_3}\right)\left(\frac{K_2}{K_1}-\frac{K_2}{2K_3}\right)\right];$$

$$C=+\frac{1}{2}\left(\frac{K_2}{K_3}\right)^2\left[\frac{1}{2}\left(1+\frac{\alpha}{\beta}\right)^2+\frac{\alpha}{\beta}\right]-\left(\frac{2\alpha K_2}{\beta K_1}\right)\left(\frac{K_3}{2p'}\right)\left\{\frac{\alpha}{\beta}\frac{K_2}{K_3}\left[1-\frac{K_2}{2K_3}\left(\frac{3}{2}+\frac{\alpha}{\beta}\right)+\frac{\alpha K_2}{2\beta K_1}\right]\right.$$
$$\left.+\frac{K_2}{K_3}\left[\frac{\alpha}{\beta}\left(\frac{K_2}{K_1}-\frac{K_2}{2K_3}\right)+\frac{1}{2}\left(1-\frac{K_2}{K_3}-\frac{\alpha K_2}{2\beta K_3}\right)\right]\right\};$$

$$D=-\frac{1}{2}\left(1+\frac{\alpha}{\beta}\right)\left(\frac{K_2}{K_3}\right)^2\left(1+\frac{K_3}{2p'}\right);$$

$$E=+\frac{1}{4}\left(\frac{K_2}{K_3}\right)^2.$$

For a given T and p' the biquadratic can be solved for z; x, y and v follow from equations (26), (27), and (23). From T and v, p_A follows, and hence p from (25). The x, y, z factors fix the composition of air and hence the component energies. The E_2 value is then the sum of these. This method is a very laborious one as it entails initially the evaluation of the coefficients A, B, etc., which in itself is not easy, and then the solution of the biquadratic for a large number of p' points.

An alternative procedure was developed which proved much simpler. Fixing the temperature T_2, a value of z is chosen which, with the aid of (26), (27), (28), (24) and (25), yields a p_2 value in the neighbourhood of the shock wave conditions. It is here that the first solution is of use. With a little practice, choice of a z value at any T_2 to give an approximate p_2 becomes an easy matter. E_2 follows from x, y, z and hence ΔE. ΔE is also computed from equation (18). The correct "shock wave z" is the one which makes E equal by the two methods. The correct U, u, p_2, v_2, etc., for the given T_2 then follow from the shock wave equations.

At 3000° K. the nitrogen component has not commenced to dissociate appreciably, so that no nitric oxide has been formed at this stage. At 4000° K. air at shock wave conditions contains about 4·4%, at 5000° K. about 4·8% by weight of NO. The percentage then slowly decreases with increasing temperature until at 14 000° K. the content has fallen to 0·27%.

The results of these computations are shown in table 5. The temperature T_2 in this case is the independent variable.

At temperatures below 4000° K. the contribution made by the energy of dissociation to the internal energy is small, but at 14000° K. it is as much as 60%. It is computed by assessing the energy liberated by the initial dissociation of the nitrogen and oxygen molecules and subtracting a certain amount due to the recombination of some of the atoms to form nitric oxide.

Table 5. Shock wave conditions in air

T_2(° K.)	4,000	6,000	8,000	10,000	14,000
x	0·1571	0·7298	0·9257	0·9678	0·9917
y	0·000092	0·0092	0·0885	0·2968	0·7809
z	0·1025	0·0875	0·0430	0·0226	0·0061
% mass in air of					
NO	4·46	3·81	1·87	0·98	0·27
O	3·64	16·93	21·48	22·45	23·01
O_2	17·18	4·24	0·73	0·22	0·05
N	0·01	0·70	6·68	22·41	58·96
N_2	73·41	73·02	67·94	52·64	16·41
A	1·30	1·30	1·30	1·30	1·30
ΔE (cal/gm.)	1040	2024	3090	4790	8540
p_2 (atmos.)	127	236	366	561	1003
v_2 (cm³/gm.)	91·7	81·8	78·4	72·5	71·9
U (m/sec.)	3350	4540	5645	6965	9310
u (m/sec.)	2950	4060	5075	6320	8445

In table 6 the shock wave variables are given with the pressure as the independent variable.

Table 6

p_2 (atmos.)	200	400	600	800	1,000
U (m/sec.)	4,180	5,890	7,200	8,310	9,310
u (m/sec.)	3,750	5,300	6,550	7,560	8,450
T_2(° K.)	5,270	8,420	10,300	12,000	14,000
E (cal/gm.)	1,730	3,280	5,100	6,670	8,420
v_2 (cm³/gm).	82·7	76·8	71·8	69·9	71·8

The U and u values are not far removed from those given in table 4 as the first solution. The T_2 figures, however, have changed a good deal, and at $p_2=500$ atmospheres T_2 has increased by 1480° K. The ΔE values are still in close proximity and show a practically linear dependence on p_2. There is now a minimum value of v_2 (about 70 cm³) at $p_2=800$ atmospheres instead of at 600 atmospheres.

§7. NORMAL IMPACT OF A SHOCK WAVE ON A RIGID WALL

The pressure experienced by a rigid wall when a shock wave strikes it is many times greater than the shock wave pressure p_2. The simple theory is quite adequate for low values of pressure, but for the higher values calculation becomes difficult. The method makes use of the fact that the air behind the returning shock wave must be at rest. Thus, if p_2 and v_2 are the pressure and volume of unit mass of air before impact, p_3 and v_3 similar quantities after impact, u the mass velocity produced immediately in the rear of the shock wave by p_2 and v_2, then

$$v_3=v_2-u^2/(p_3-p_2). \qquad \ldots\ldots(29)$$

Outside the "dissociation region" solution is easy. Assume a value for p_3, then equation (29) gives v_3 and hence T_3 by the perfect-gas laws. The internal energy of air function, $E(p, T)$, has to be plotted as a function of temperature for equal values of p_3—considerably greater in magnitude than was required for p_2. Interpolation yields $(E_3 - E_2)$ where E_2 and E_3 are the energies of air at the shock wave before and after impact. If these values agree with that obtained from the formula

$$E_3 - E_2 = \tfrac{1}{2}(p_3 + p_2)(v_2 - v_3), \qquad \ldots\ldots(30)$$

then the p_3 chosen is the correct one. Normally p_3 has to be readjusted so that this requirement is satisfied.

Inside the "dissociation zone" solution is more difficult, as $v_3(p, T)$ has to be plotted for equal pressures (figure 8). The steps are :

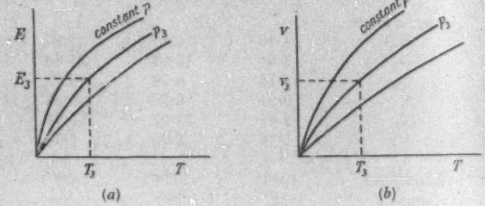


Figure 8.

(i) Assume a certain p_3 value, equation (29) gives v_3; (ii) plot 8 (*b*) gives the associated T_3; (iii) plot 8 (*a*) gives E_3; (iv) hence $E_3 - E_2$(known) $= \Delta E$; (v) compare with value given by equation (30); (vi) adjust p_3 to obtain agreement.

In the absence of NO and argon, and with the old value of D for nitrogen, the following values for p_3/p_2 have been evaluated.

Table 7

p (atmos.)	15	30	50	65	100	200	300	375	500	800	1000
p_3/p_2	5·87	6·83	7·8	8·2	8·9	10·3	12·8	13·8	14·0	12·3	10·8

It is unlikely that correction for formation of NO, etc., will affect these values to any appreciable extent.

§8. EFFECT OF IONIZATION

The method used for investigating the possible changes in the internal energy, equilibrium constants and shock wave properties, due to ionization in the earth's surface atmosphere due to the very high temperatures considered in this paper, is similar to that for evaluating the "electronic" contributions. Possible ionic formations are accounted for by a similar addition to the partition functions and yield contributions as follows:

Ionic contribution to the internal energy

$$E_i = \{i_1 I_1 \exp(-I_1/\boldsymbol{k}T)\}/\{g_0 + i_1 \exp(-I_1/\boldsymbol{k}T)\}; \qquad \ldots\ldots(31)$$

contribution to the previous $\ln K$ formula.

$$\ln\left\{\frac{\left[1+\frac{i_1^{\mathrm{A}}}{g_0^{\mathrm{A}}}\left(\exp-\frac{I_1^{\mathrm{A}}}{\boldsymbol{k}T}\right)\right]\left[1+\frac{i_1^{\mathrm{B}}}{g_0^{\mathrm{B}}}\left(\exp-\frac{I_1^{\mathrm{B}}}{\boldsymbol{k}T}\right)\right]}{\left[1+\frac{i_1^{\mathrm{AB}}}{g_0^{\mathrm{AB}}}\left(\exp-I_1^{\mathrm{AB}}/\boldsymbol{k}T\right)\right]}\right\}, \qquad \ldots\ldots(32)$$

where i_1 is the statistical weight of atoms (or molecules) in the main ionized state and I_1 the energy difference from ground state to ionized state.

These equations cover the transformations

$$N_2 = N_2^+ + e,\ O_2 = O_2^+ + e,\ O = O^+ + e,\ N = N^+ + e. \qquad \ldots\ldots(33)$$

The spectroscopic data required are taken from *Molecular Spectra and Molecular Structure of Diatomic Molecules* by Herzberg, and *Atomic Energy States* by Bocher and Goudsmidt.

	Electronic potentials (ev.)	State	Statistical weights
O^+	13·60	^4S	4
N^+	14·54	^3P	1
O_2^+	12·20	$^2\Pi^g$	2
N_2^+	15·58	$^2\Sigma^+$	2

These electronic potentials are very considerably higher than the excited levels of the un-ionized gases. Indeed, substitution of these numerical values in equations (31) and (32) shows immediately that ionic formation at high temperatures has a negligibly small effect on the internal energies and K values. It does not, therefore, affect the shock wave properties.

For temperatures greater than 16000° K. ionization plays an increasing part; the calculations to allow for this are described by Dr. Kynch in his papers in the "Atomic Energy" series. The problem at these very high temperatures is eased by complete dissociation.

ACKNOWLEDGMENTS

The author wishes to thank the Chief Scientific Adviser, Ministry of Home Security, for permission to publish this paper, and Dr. W. G. Penney for many helpful discussions.

REFERENCE

Lewis, B., and von Elbe, G., 1938, *Combustion, Flames and Explosion of Gases* (Cambridge: University Press), pp. 239–244.

DISCUSSION.

Dr. A. H. Davis. The present paper deals with the relations which exist between pressure, wave velocity, particle velocity, etc., whereby the observations of one quantity at a point—say of the velocity of the wave—makes it possible to infer the magnitude of others—say of the pressure, temperature cycle and so on. It does not deal with the propagation of shock waves through the atmosphere. Moreover, whilst theoretical work has been conducted on the propagation of sounds of "small" amplitude through the atmosphere, around obstacles and so on, I believe that there is not yet any theory of the propagation of finite shock waves under similar conditions, and consequently an interesting theoretical field lies open.

The genesis of a shock wave may be observed by exploding an explosive mixture of hydrogen and oxygen contained in an india-rubber balloon of about 6 cubic feet in volume. (The operations of filling the balloon, first partially with one gas, and then finally with the other, must be conducted by remote control methods, as the balloon is liable to detonate when handled under some atmospheric conditions.) Observations of the blast pressure at distances of some 3 ft. to 10 ft. from the explosion reveal a wave which is first rather nondescript, but which develops into a shock wave after travelling a few feet.

Mr. A. Warren. I should like to draw attention to some unusual features in the shape of the shock waves recorded near the ground with small high explosive charges.

The measurements were made using an experimental technique recently described by Grime and Sheard (*Proc. Roy. Soc.* A, 1946, **187**, 357) up to a distance at which the calculated

velocity of the shock wave is within $\frac{1}{4}\%$ of the normal velocity of sound, and under these circumstances it was anticipated that any change in the wave form of the shock pulse would be insignificant.

It was found that up to a certain distance a wave-form characteristic of normal shock propagation was recorded, but that subsequently a change of waveform occurred ; a secondary pressure peak commenced to form behind the shock front and the two peaks tended to separate. Since the maximum excess pressure in the initial shock front decayed at an abnormally high rate, the maximum excess pressure in the secondary peak soon exceeded that in the shock front, and in some instances the initial shock front almost disappeared at the greater distances.

It was observed that the critical distance beyond which the secondary peak pressure exceeded that in the shock front appeared to be greater when the propagation losses in the wave front were reduced. For example, it was greater (*a*) in the direction of the prevailing wind, (*b*) for gauges at (say) 3 feet above ground as compared with similar measurements very near the ground, and (*c*) if the surface of the ground between the charge and gauge was covered with sheets of smooth material, e.g. strawboard with improved reflection characteristics.

It was known from the recent series of papers by Friedlander (" Diffraction of sound pulse", *Proc. Roy. Soc.* A, 1946, **186**, 322) that after diffraction from obstacles of various types a shock wave was likely to lose sharpness and become broadened. A similar effect was anticipated in an unpublished mathematical treatment by Prof. Sir. G. I. Taylor dealing with the effects due to turbulence, and to the gradient of wind velocity near the surface of the ground.

It appeared from the records already obtained that modified wave-forms very similar to those to be anticipated on theoretical grounds from the effects of diffraction or turbulence were observed, but that in addition, under certain circumstances, this modified wave-form could co-exist with an initial shock front of characteristic shape at an appropriate time separation.

Mr. A. R. Bryant. I should like to ask what the temperature would be if an explosive were surrounded by an envelope of argon instead of air. One would expect this to be much higher owing to argon having a higher value of γ and owing to absence of dissociation. With regard to the results of the shock wave calculations could the author state how sensitive the quantities behind the shock front are to the assumption that thermodynamic equilibrium is or is not attained ?

Author's reply. I have read with interest the remarks of Dr. A. H. Davis and Mr. A. Warren. An approximate solution of the problems of hydrodynamical flow behind a shock wave set up in air as a result of an explosion from a spherical charge was, in fact, attempted in 1941 by Dr. W. G. Penney and myself (unpublished). At that time the adiabatics of air at high temperature and pressure had not been evaluated, and γ was taken to be equal to 1·403. This assumption was likely to introduce large numerical errors, but the results showed the presence of a secondary pressure peak, as mentioned by Mr. Warren. The problem was investigated during 1943–45 in the U.S.A., but I am not aware if any of this work has yet been published.

In reply to Mr. Bryant's first question, I agree that the temperature would be higher if the explosive were surrounded by argon of the same density as air in the normal atmosphere, the internal energy of argon per gramme at a given temperature being less than that of air ; there is no rotational, vibrational energy, and no dissociation to be allowed for.

With regard to Mr. Bryant's second question, the shock wave velocity and pressure are probably very little affected by the assumption of thermodynamic equilibrium. The actual temperature, however, will be higher than those obtained on this assumption. In the paper it is assumed that the energy is taken up in the translational, rotational and vibrational forms, distributed according to the equations of statistical mechanics. A given volume of air receives a definite quantity of energy from the " shock wave ", and it is probable that, due to the nature of the " shock ", rather more energy is taken up into the translational type and less into the rotational and vibrational types than given by the theory. Hence the temperature would be higher than expected on the assumption of thermodynamic equilibrium.

The Reflection of a Plane Shock Wave at a Gaseous Interface

By STEWART PATERSON
Imperial Chemical Industries Ltd, Explosives Division, Stevenston, Scotland

MS. received 18 *February* 1948

ABSTRACT. A plane shock wave falls normally on the interface between two ideal gases of molecular weights M, m and constant ratios Γ, γ of specific heats. As a rule, the character of the reflected wave then depends only on the ratio $\Gamma M/\gamma m$. In exceptional cases, however, the type of reflection may change at a critical value of the incident intensity. An example of this is given.

Two perfect gases with specific volumes V_0 and v_0 and constant ratios Γ, γ of specific heats are at rest and separated by an ideally thin and light partition. The pressure is p_0 and temperature T_0 in each gas. A plane shock wave S_1, the steady conditions behind which are indicated by suffix 1, passes through the first gas and falls normally on the interface.

A wave S_2 will then be transmitted through the second gas; we denote conditions behind S_2 by a suffix 2. These conditions are defined by the familiar Rankine-Hugoniot relation if S_2 is a shock wave, or by the Riemann relation for a progressive wave if S_2 is a rarefaction. Thus, if W represents in general the mass velocity behind a wave (measured in the direction of propagation of S_1), and π the pressure ratio across it, so that $\pi_1 \equiv p_1/p_0$, $\pi_2 \equiv p_2/p_0$, we have for S_1

$$W_1 = \sqrt{\{(\Lambda-1)p_0V_0\}}(\pi_1-1)/\sqrt{(\Lambda\pi_1+1)}, \qquad \ldots\ldots(1)$$

and for S_2

$$W_2 = \begin{cases} \sqrt{\{(\lambda-1)p_0v_0\}}(\pi_2-1)/\sqrt{(\lambda\pi_2+1)}, & \text{if } \pi_2>1 \\ -\sqrt{\{(\lambda^2-1)p_0v_0\}}[1-\pi_2^{1/(\lambda+1)}], & \text{if } \pi_2<1 \end{cases} \qquad \ldots\ldots(2)$$

where $\Lambda \equiv (\Gamma+1)/(\Gamma-1)$ and $\lambda \equiv (\gamma+1)/(\gamma-1)$. It is of course physically obvious, and will be shown below, that S_2 must be a shock wave ($\pi_2>1$) and not a rarefaction wave ($\pi_2<1$). However, for the moment we need not assume this.

A reflected wave S_3 will also arise in the first gas; we denote conditions behind by a suffix 3, so that $\pi_3 \equiv p_3/p_1$. S_3 may be either a shock wave ($\pi_3>1$) or a rarefaction wave ($\pi_3<1$), and the mass velocity behind it is

$$W_3 = \begin{cases} W_1 - \sqrt{\{(\Lambda-1)p_1V_1\}}(\pi_3-1)/\sqrt{\{\Lambda\pi_3+1\}}, & \text{if } \pi_3>1 \\ W_1 + \sqrt{\{(\Lambda^2-1)p_1V_1\}}[1-\pi_3^{1/(\Lambda+1)}], & \text{if } \pi_3<1. \end{cases} \qquad \ldots\ldots(3)$$

Continuity of velocity and pressure across the interface requires, however, that $W_2 = W_3$ and also $p_2 = p_3$, so that $\pi_2 = \pi_3\pi_1$. Making these substitutions in (2) and using (3) and (1), we are in a position to determine π_3, and so all the properties of the reflected and transmitted waves. The procedure is now well known.

It may easily be shown that W_2, as given by (2) with $\pi_2 = \pi_3\pi_1$, is a monotonic function of π_3 which increases continuously through zero as π_3 increases through $1/\pi_1$. In the same way, W_3 is a monotonic function of π_3 which decreases continuously through W_1 as π_3 increases through unity. It therefore appears at once, since $W_1>0$,

(*a*) that π_3 cannot be less than $1/\pi_1$; in other words, that π_2 must always exceed 1, so that the transmitted wave is necessarily a shock wave;

(*b*) that π_3 may be either greater than or less than 1, so that the reflected wave may be of either type, according to circumstances.

These properties are illustrated in figures 1 and 2 by the case of a 2-atm. shock wave proceeding from oxygen to argon and vice versa, both gases being initially at 1 atm. and 273° K.

The condition that S_3 be a shock wave is that the two curves intersect at a point for which $\pi_3>1$; but in view of the above remarks concerning the course of $W_2(\pi_3)$ and $W_3(\pi_3)$, this condition can be replaced by the analytically much simpler one that $W_3(\pi_3)>W_2(\pi_3)$ when $\pi_3=1$; in other words, that

$$(\Lambda-1)V_0/(\Lambda\pi_1+1)>(\lambda-1)v_0/(\lambda\pi_1+1), \qquad \ldots\ldots(4)$$

which may be written

$$(\pi_1-1)[(\Gamma+1)\Delta_0-(\gamma+1)\delta_0]<2[\gamma\delta_0-\Gamma\Delta_0], \qquad \ldots\ldots(5)$$

where $\Delta_0 V_0 \equiv 1 \equiv \delta_0 v_0$.

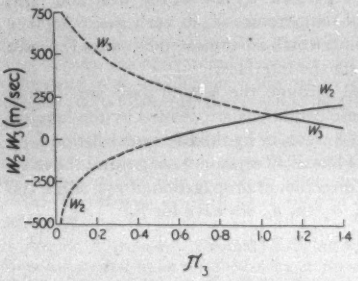


Figure 1. Reflection of a 2-atm. shock wave in oxygen by argon. $p_0=1$ atm.; $T_0=273°$ K; W_2, W_3 ≡ mass velocities behind transmitted and reflected waves; π_3 ≡ pressure ratio across reflected wave, which is seen to be a shock.

——————— Shock
– – – – – – – – – – – – – Rarefaction

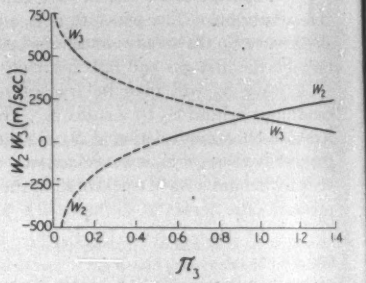


Figure 2. Reflection of a 2-atm. shock wave in argon by oxygen. $p_0=1$ atm.; $T_0=273°$ K.; W_2, W_3 ≡ mass velocities behind transmitted and reflected waves; π_3 ≡ pressure ratio across reflected wave, which is seen to be a rarefaction.

——————— Shock
– – – – – – – – – – – – – Rarefaction

Four alternatives now arise:

(*a*) $\Gamma\Delta_0<\gamma\delta_0$ *and* $(\Gamma+1)\Delta_0<(\gamma+1)\delta_0$.

Then (5) is satisfied irrespective of $\pi_1(>1)$, and the reflected wave is always a shock wave.

(*b*) $\Gamma\Delta_0>\gamma\delta_0$ *and* $(\Gamma+1)\Delta_0>(\gamma+1)\delta_0$.

Then (5) cannot be satisfied for any $\pi_1(>1)$, and the reflected wave is always a rarefaction wave.

(*c*) $\Gamma\Delta_0<\gamma\delta_0$ *but* $(\Gamma+1)\Delta_0>(\gamma+1)\delta_0$.

Then (5) is satisfied for $\pi_1<\pi_1'$, but not for $\pi_1>\pi_1'$, where

$$\pi_1' \equiv \frac{(\gamma-1)\delta_0-(\Gamma-1)\Delta_0}{(\Gamma+1)\Delta_0-(\gamma+1)\delta_0}. \qquad \ldots\ldots(6)$$

Thus, for $\pi_1<\pi_1'$, S_3 is a shock wave, but for $\pi_1>\pi_1'$, S_3 is a rarefaction wave.

(*d*) $\Gamma\Delta_0>\gamma\delta_0$ *but* $(\Gamma+1)\Delta_0<(\gamma+1)\delta_0$.

Then, for $\pi_1<\pi_1'$, S_3 is a rarefaction wave, but for $\pi_1>\pi_1'$, S_3 is a shock wave.

Cases (*a*) and (*b*) are evidently reciprocal. Thus they show that if shock waves in the first gas are always reflected as shock waves by the second, shock waves in the second will always be reflected as rarefaction waves by the first, and vice versa. This is physically reasonable. Cases (*c*) and (*d*) are reciprocal in a similar sense, but the existence of a critical incident pressure ratio π_1', at which the reflection changes type, is more surprising. It remains to be seen whether this effect can be realized in real gases. The necessary condition may be written, for convenience,

$$\Gamma M \lesseqgtr \gamma m, \qquad (\Gamma+1)M \gtreqless (\gamma+1)m, \qquad \ldots\ldots(7)$$

where M, m denote molecular weights. Examination of tabulated specific heats and molecular weights shows at once that (7) is not usually satisfied, so that cases (*a*) and (*b*) form the rule. However, the possibility of satisfying (7) is not excluded if Γ and γ are sufficiently different. Suppose, for example, that $\Gamma=1{\cdot}1$ and $\gamma=1{\cdot}667$. Then (7) requires that $1{\cdot}27m<M<1{\cdot}515m$. It appears in fact that (7) is satisfied by the following pairs of gases: neon with acetylene, argon with butane vapour or methyl chloride vapour, for which approximate constants at 1 atm. and 273° K. and corresponding values of π_1' are given in the following table:

Gas	Γ	Δ	$\Gamma\Delta$	$(\Gamma+1)\Delta$	π_1'
Neon (Ne)	1·642	0·966 (g/l.)	1·586	2·553	(with Ne)
Acetylene	1.28	1·190	1·52	2·71	1·78
Argon (A)	1·667	1·781	2·964	4·750	(with A)
Butane vapour	1·11	2·599	2·88	5·48	1·51
Methyl chloride vapour	1·279	2·238	2·862	5·10	1·60

Calculations for the vapours are rather unreliable, in view of the high critical temperatures and correspondingly large departures from the ideal gas laws. For neon and actylene, however, the effect is probably authentic, and could perhaps be detected in Schlieren photographs by the reversal of the sign of π_3-1 at π_1'; the wave- and mass-velocities do not, of course, change sign. The properties of transmitted and reflected waves in this system up to $\pi_1=2{\cdot}5$ are shown in figure 3.

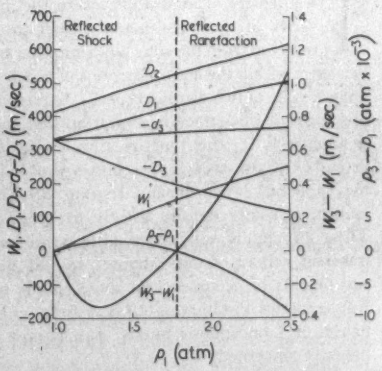


Figure 3. Reflection of weak shock waves in acetylene by neon. $p_0=1$ atm.; $T_0=273°$ K.; p_1, p_3 = pressures, and W_1, W_3 mass velocities, behind incident wave and between transmitted and reflected waves respectively; D_1, D_2, D_3 = absolute velocities of incident, transmitted and reflected waves; d_3 = velocity of reflected wave relative to gas behind incident wave. Velocities are positive when in the direction of incident wave.

For the great majority of gaseous systems (7) fails, and the condition for a reflected shock is then simply $\Gamma M<\gamma m$, which may also be written $A_0M<a_0m$ or

$$\left|\left(\frac{\partial V_0}{\partial p_0}\right)_S\right|<\left|\left(\frac{\partial v_0}{\partial p_0}\right)_S\right|,$$

where A_0, a_0 denote velocities of sound, and S entropy.

A Study of Transient Radar Echoes from the Ionosphere

By E. EASTWOOD and K. A. MERCER

Communicated by E. V. Appleton ; MS. received 21 *February* 1947 *and in amended form* 22 *November* 1947; *read* 31 *January* 1947

ABSTRACT. The rate of occurrence of transient bursts of ionization within the lower E region of the ionosphere has been systematically measured during the period January 1945 to July 1946.

It has been found that the activity varies both diurnally and seasonally and in such a manner as to lend support to the meteor theory of burst formation. Some measure of solar control of the rate of burst occurrence has also been detected ; this effect may be explained in terms of the normal diurnal E region changes in density of ionization. From experiments performed during the solar eclipse of July 1945, it is concluded that the sun does not emit a burst-producing radiation.

Analysis of the observations suggests that the greater proportion of the bursts are created within a thin layer located at a height of 86 km. ; the distribution of the bursts within this layer has proved to be uniform over wide areas and no latitude effect has been detected. It is established that the rate of incidence of bursts which present echoing areas between A and $A+dA$ m² to a radio wave of frequency ν may be expressed in the form $CdA/\nu^3A^{3/2}$, with C constant.

§1. INTRODUCTION

The density of ionization within the E region of the ionosphere normally varies in a regular manner with the intensity of the ultra-violet light received from the sun, but a number of workers have shown the temporary existence within the layer of regions of much higher density. These disturbances appear to be of two distinct types. Appleton and Naismith (1935, 1940) describe one form of this sporadic ionization that has come to be known as "abnormal or intense E"; this type usually persists for minutes or even hours and may extend over large areas. A second type of E-layer disturbance was also recognized by Appleton, Naismith and Ingram (1937) in the form of localized bursts of ionization whose density greatly exceeded that characteristic of abnormal E. These ion clouds were observed to occur both by day and by night and yielded transient echoes of only about a second duration. The reflection coefficients and heights of occurrence of these clouds were later studied in more detail by Appleton and Piddington (1938). It is with this second type of E-layer disturbance, or "ionospheric burst" as it is now commonly termed, that the present paper is concerned.

The problem of E-layer disturbances was examined by Eckersley (1940) from the point of view of radio communication effects and he has shown that the presence of signals within the skip zone is attributable to scattering from ionospheric "clouds". The rate of occurrence of the clouds was found to follow a diurnal cycle, while the dependence of the intensity of the scattered radiation upon the wavelength was in agreement with a fourth power law previously suggested by him (1932).

The nature of the ionizing agent responsible for these E-layer perturbations is a question which has given rise to speculation, but much of the evidence now

points towards the validity of the meteor theory. Nagaoka (1929) was the first to consider the possible influence of meteors on radio transmission and suggested that disturbed ionospheric conditions might arise from the removal of electrons by a process of attachment to particles of meteoric dust. Skellett (1932, 1935) disagreed with this theory and, supported by experimental work performed in collaboration with Schafer and Goodall (1932), showed that impact ionization by meteors was a more likely explanation of the observed phenomena. Pierce (1938) also found that sudden increases in the abnormal E type of ionization could be explained by the passage overhead of a single large meteor, while certain experiments reported by Eckersley (1937) on the occurrence of scattering clouds were considered by Skellett (1938) to be consistent with the meteor theory of their formation. Further convincing evidence in support of the meteor hypothesis has recently been provided by Appleton and Naismith (1946), and by Hey and Stewart (1946), but that certain difficulties still remain in applying the theory is shown by the work of Eckersley and Farmer (1945) on the rapid variation of the direction of arrival and state of polarization of the waves scattered from the bursts.

The aim of the present investigation has been to gather further information upon the characteristics of ionospheric bursts by making a systematic study of the time variation of burst activity. Observations were also made during the total solar eclipse of July 1945 to detect any influence which the sun might exert upon the phenomenon of burst occurrence.

§ 2. EXPERIMENTAL METHODS

The "pulse-echo" technique has been used in the present work for the detection and location of ionospheric bursts. The experiments have been conducted at a number of Royal Air Force radar stations each of which was equipped with a high power, pulsed radio transmitter and a sensitive receiver. The transmitter aerial consisted of a vertical stack of horizontal, half-wave dipoles and was situated within 200 metres of a receiver aerial with dipole elements parallel to and at the same mean height as those of the transmitter. The radio-frequencies employed ranged from 20 to 45 Mc/sec. but the pulse recurrence frequency was 25 pulses/sec. at all the stations.

The echo signal returned by the burst was recorded photographically in the usual ($P't$) type of record. With the aerial system described above most of the energy is radiated at comparatively low angles of elevation; a method of oblique incidence has therefore been used, and the P' of the records represents not the usual equivalent height of the ionospheric reflecting layer but the slant range of the ionospheric burst. The magnitudes of the echoes were also recorded photographically using a ciné camera which was operated without its shutter. It was found that successive sweeps on the cathode-ray display tube were registered as separate parallel traces on the film so that the echo amplitudes could easily be measured from the trace deflections. This simple technique permitted the rates of growth and decay of particular responses to be studied.

§ 3. DIURNAL AND SEASONAL VARIATIONS

In order to investigate diurnal and seasonal changes in the rate of burst occurrence observations were made from two identical stations at Bawdsey, Suffolk, on a frequency of 22·69 Mc/sec., a pulse width of 8 μsec. and a power

of 1 Mw. The hourly rate of occurrence of bursts formed within a selected slant range interval was measured over periods of twenty-four hours. The observations covered the period January 1945–July 1946, with the 24-hour sequence of observations taking place twice each week.

Every precaution was taken to ensure constant performance of the equipment but output power variations did occur and, in order to apply appropriate correction, the relationship between the number of bursts observed and the power radiated was separately investigated. This work established the result that, for a given system and range interval, the number of bursts observed was directly proportional to the square root of the power. From this result it was now possible to reduce the average measured power over an interval to the standard power of 1 Mw. and to correct the burst rate accordingly.

The number of bursts recorded during the successive hours of a day's run clearly showed the presence of a diurnal cycle of burst activity. These daily curves possessed certain features in common but marked changes also occurred from day to day and curves of mean values for each month were therefore plotted (figures 1 and 2). The curves of figure 1 show relative monthly changes in the daily cycle; the absolute levels of activity are not comparable since the same aerial system was not used from month to month. The curves of figure 2, for which the same aerial system was used, are however, strictly comparable both for total activity and for relative monthly changes.

Figures 1 and 2 show clearly the presence of a diurnal cycle of burst activity and also show that the form of the daily cycle changes with the time of the year. The diurnal cycle is characterized by three main features—a pronounced minimum in the late afternoon, and maxima in the early and late morning. It will be seen that the minimum of the daily curve is the most clearly defined and persistent feature of all the histograms and that its time of occurrence is approximately constant and independent of the time of sunset. The two maxima of the daily curve occur at

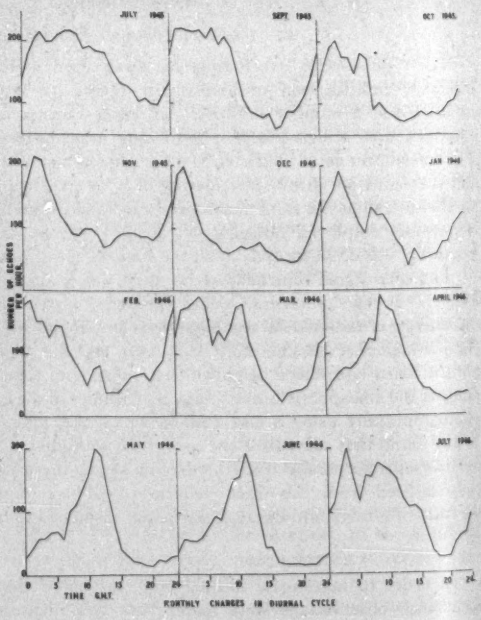


Figure 1.

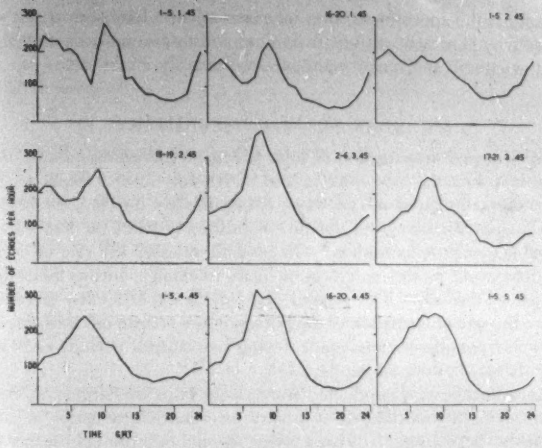


Figure 2.

approximately 0400 and 1200 U.T., and do not appear to vary with time of sunrise and sunset through the year. The relative amplitudes of the two maxima are, however, dependent upon the season of the year; in particular, the early morning peak tends to be associated with the period of the winter solstice while the late morning peak becomes the dominant feature of the curve about the time of the summer solstice. This is seen in the middle column of curves in figure 1 : December 1945 and June 1946 show pronounced early and late morning peaks respectively, while September 1945 and March 1946 have transitional types of curves with both peaks present and of roughly the same amplitudes.

The change in total volume of daily burst activity throughout the year was studied by taking observations on convenient days when the same aerial system could be used. The total number of echoes during a 24-hour period observed within the range interval 160–320 km. was determined and the average number of responses per hour extracted. Figure 3 shows these average daily figures

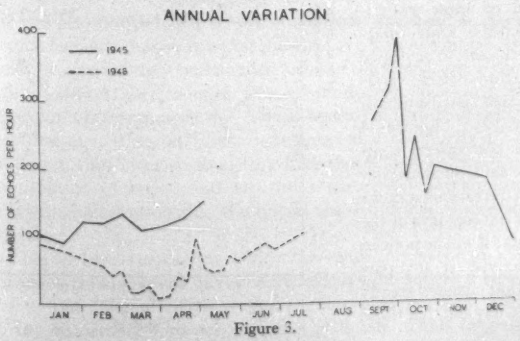


Figure 3.

plotted against the appropriate dates of observation. It is seen that a seasonal cycle of activity is present with its maximum and minimum occurring about the times of the autumn and spring equinoxes respectively.

§4. SOLAR ECLIPSE EXPERIMENTS

An experimental investigation of solar effects on ionospheric burst formation was possible during the total solar eclipse of 9th July 1945. Such solar effects might be expected to arise in two ways, either directly due to the emission of a burst-producing radiation from the sun or indirectly from the sun's control of the ionized state of the ionosphere. In both these cases the rate of occurrence of the bursts would be the feature most likely to change during the course of a total eclipse of the sun. The general plan of the experiments, therefore, was to compare the rate of formation of the bursts on the eclipse day with the average rate obtained from observations made during two control periods each of seven days, on either side of the eclipse day.

The experiments were conducted from two groups of stations in the north of Scotland, lying within the belt of totality for a particle radiation at the range of velocities to be expected. The activity throughout the eclipse period was measured in terms of the number of echoes observed per half hour and was compared with the mean activity derived from a grouping of the corresponding periods of the control days. The amplitudes and durations of echoes observed during the eclipse were similarly compared with the averages derived from the control periods. This procedure was followed at all the stations of table 1.

A careful statistical analysis of all the results obtained failed to reveal any significant departure of ionospheric burst characteristics observed during the period of the eclipse from the averages established by the control periods. It may be concluded therefore, that the sun is not a source of burst-producing particles. However, the great volume of data obtained in the course of these experiments permitted some interesting conclusions to be drawn on the questions of time variation and space distribution of burst activity. Figure 4 presents a convenient summary of the daily burst activity recorded at the various stations during the sixteen days of the experiments; it will be noticed that the activity at any particular station changes greatly from day to day and that these

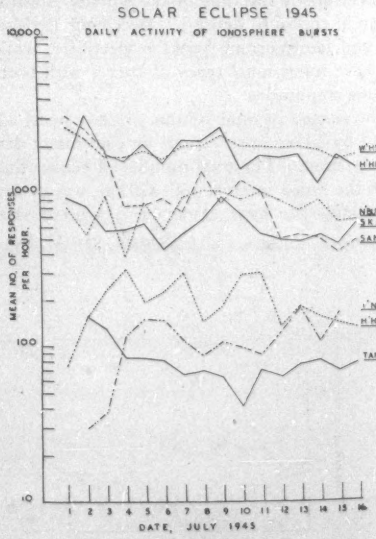


Figure 4.

changes occur in an apparently random manner at the different stations. The large differences between the mean activity levels of the stations taken over the whole duration of the observations are particularly noteworthy and will be discussed in more detail later.

§ 5. RANGE DISTRIBUTION OF BURSTS

From the $(P'\,t)$ records obtained during the eclipse experiments a histogram was derived for each of the stations relating the number of echoes in successive 16 km. intervals of range with the mean ranges of these intervals. It was found that every station possessed a characteristic range-distribution histogram. Similarly, each of the various aerial systems used at Bawdsey was found to have a characteristic histogram; figure 5 shows a typical histogram secured from a two-stack receiver aerial system combined with an eight-stack transmitter array, both aerials centred on a mean height above sea-level of 255 ft. and operated on a frequency of 22·69 Mc/sec. Figure 6 shows the distribution corresponding to a single dipole receiver aerial and a single dipole transmitter, each at a height of one-quarter wavelength above the ground.

In all these experiments a sharp cut-off in range was observed at about 80 km., in agreement with the accepted view that the bursts are formed within the E

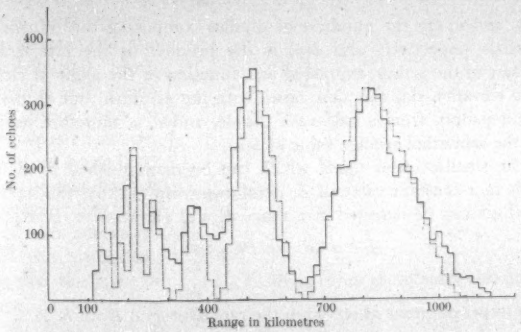


Figure 5. Range distribution.
——— Bawdsey. ..... Theoretical.

layer, but it was also noticed that echoes occasionally appeared at less than this range. Detailed analysis has shown that the form of a station's range-distribution histogram may be convincingly explained in terms of the interaction of the station's characteristic radiation pattern with a burst-producing layer situated within the lower E region of the ionosphere. The sharpness of certain of the histogram minima, e.g. that at 650 km. in figure 5, suggests that the bursts are probably confined to a very thin layer so that, to a first approximation, we may neglect a height distribution of bursts and seek for a theoretical development of the histograms on the assumption that the bursts are distributed uniformly over a spherical surface.

The number of bursts occurring within the range interval R to $R+dR$ and having equivalent echoing areas between A and $A+dA$ square metres will be proportional to the zone of the ionosphere intercepted between these ranges and also to the occurrence rate per unit area of bursts of this magnitude, say

$n\,dA$. If h is the height of the burst-producing layer and ρ be the radius of the earth, then the area of the zone is $2\pi(1+h/\rho)R\,dR$, and we may therefore write :

Number of bursts observed in the range interval R to $R+dR$ having echoing areas between A and $A+dA$: $\propto n\,R\,dA\,dR$, ...(1)

i.e., total number of bursts observed in the range interval R_1 to R_2: $\propto \int_{R_1}^{R_2}\int_{A_L}^{\infty} nR\,dA\,dR$. .....(2)

In this expression A_L is the smallest echoing area which can be detected by the given equipment at the range R and it will be shown later (equation (9)) that

$$\int_{A_L}^{\infty} n\,dA \propto 1/\sqrt{A_L},$$

so that (2) may be rewritten :

Total number of bursts observed in the interval R_1 to R_2: $\propto \int_{R_1}^{R_2} R\,A_L^{-\frac{1}{2}}dR$. .....(3)

Now a burst of echoing area Am^2 formed at a range R from a station radiating a power of P kilowatts on a wavelength of λ m. will return an echo whose signal/noise ratio is Z, where Z is given by

$$Z \propto \lambda R^{-2}\sqrt{(APn_tn_r)}\phi(\alpha), \qquad .....(4)$$

in which n_t and n_r are the numbers of dipoles composing the transmitter and receiver aerials respectively and $\phi(\alpha)$ is the product of the interference and grating factors of the arrays, expressed as a function of the angle of elevation α. For a given elevation this function varies with the azimuth, due to the form of the radiation pattern from a half-wave dipole, and it is therefore necessary to work with the azimuthal average value of $\phi(\alpha)$.

Since the smallest echo signal which can be distinguished from the noise corresponds to a constant value of Z, it follows from (4) that the least echoing area, A_L, which can be detected at a range R and elevation α is expressed by

$$A_L^{-\frac{1}{2}} \propto \lambda R^{-2}\sqrt{(Pn_tn_r)}\phi(\alpha). \qquad .....(5)$$

Substituting this value of A_L in (3) gives

Total number of bursts observed in the range interval R_1 to R_2

$$\propto \int_{R_1}^{R_2} \lambda R^{-1}\sqrt{(Pn_tn_r)}\phi(\alpha)dR, \qquad .....(6)$$

which may be expressed

$$\propto \int_{R_1}^{R_2} Q(R,h)dR, \qquad .....(7)$$

where h is the height of the layer of burst formation.

In order to develop the range distribution histogram for a particular system the appropriate $Q(R,h)$ was evaluated and plotted against R for a selected value of h, suitable correction being applied for the effects of atmospheric refraction. The resulting curve was integrated over successive 16 km. range intervals by use of Simpson's Rule and the areas of the consecutive blocks compared with the experimental histogram as in figure 5, where the chain lines show the theoretical histogram derived for the Bawdsey system. Theoretical histograms were similarly calculated for all the stations and for a number of selected values of h; comparison of the theoretical and experimental patterns showed that the most probable value for the height of the burst-forming layer was 86 ± 2 km. The

question of a distribution of the bursts in height was examined by a method which was an elaboration of that given above, but no better measure of agreement with experiment was obtained than that yielded by the simple assumption of a thin layer of origin. It was therefore concluded that for oblique incidence and for radio frequencies in the range 20–45 Mc/sec. the bursts may be regarded as producing their scattered waves from within a very thin layer situated at a height of 86 km.

§6. ECHOING AREAS OF BURSTS

The range of echoing areas presented by ionospheric bursts was studied by the use of a simple arrangement of two parallel, horizontal half-wave dipoles each supported one quarter wavelength above the ground and operated as the transmitter and receiver aerials respectively. The radiation pattern of this system consisted of a single vertical lobe. In order to permit the estimation of actual echoing areas as distinct from relative values only, the system was standardized from observations made during test flights with an aircraft of known echoing area.

($P't$) records of burst occurrences were secured with this system and the range distribution histogram compiled as in figure 6. The mean rate of observed occurrences per unit area of the burst layer could then be calculated at ranges corresponding to the centres of the successive blocks of the histogram; assuming that the layer of occurrence is thin, for oblique incidence and at ranges much greater than 86 km. the volumes of successive zones of this layer will be proportional to the areas intercepted on the assumed surface of occurrence situated at a height of 86 km. It is then sufficient to consider the superficial density of occurrence of the bursts on this equivalent surface.

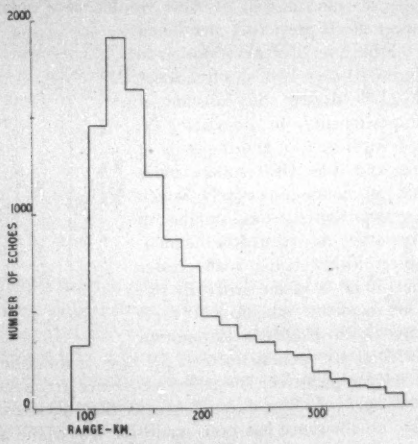


Figure 6. Range distribution.

Using the constants of the standardized aerial system the values were calculated of the smallest echoing areas that could be detected at points on the surface corresponding to the centres of the successive 16 km. zones. Let the minimum detectable echoing areas be A' and A'' for two zones for which the occurrence rate of observed echoes per unit area are N' and N'' respectively. If we assume that the distribution of bursts over the ionosphere is uniform then $(N'-N'')$ must be the rate of incidence per unit area of bursts having echoing areas between A and A''.

The number of bursts per unit area of the ionosphere occurring in unit time which present echoing areas between A and $A+dA$ we have already defined as ndA, so that the number included in the interval A' to A'' will be $\int ndA$, i.e. $(N'-N'')=\Delta N=\int_{A'}^{A''} ndA$. Now this is the area under the (n, A) curve included between the A' and A'' ordinates, and since the mean ordinate over this interval, $\Delta N/(A'-A'')$, is readily calculable, the form of the (n, A) curve itself may be traced by plotting the equivalent rectangles included between ordinates erected at abscissae corresponding to the limiting echoing areas A', A'', etc., of the successive zones. In figure 7 this method of plotting has been pursued and the straight line indicated leads immediately to the relation

$$ndA \propto A^{-3/2}dA. \qquad \ldots\ldots(8)$$

It now follows that the number of bursts per unit area per unit time, N_A say, which present echoing areas greater than A will be given by

$$N_A=\int_A^{\infty} ndA \quad \text{i.e.} \quad N_A \propto 1/\sqrt{A}. \qquad \ldots\ldots(9)$$

Separate confirmation of these results were obtained in the investigation on power effects previously mentioned.

From a set of observations made during January 1946 on a frequency of 22·69 Mc/sec. the constant of proportionality in equation (9) proved to be 643, and the units of area and time 1000 square miles and 50 hours respectively, which suggests that the total number of bursts per day created in the ionosphere with echoing areas greater than 10 m² is of the order 2×10^7. This incidence rate may be compared with Skellett's assumption (1935) of 10^9 sporadic meteors per day; it must be observed, however, that due to the comparatively low pulse recurrence frequency employed in the present work only echoes of duration greater than 0·3 sec. could be read from the films; also, no allowance has been made for a possible aspect effect such as was considered by Pierce (1938).

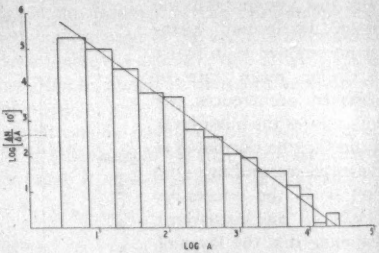


Figure 7.

No upper limit has been observed for the echoing area which it is possible for a burst to present; thus in figure 5 for example the smallest echoing area which could be detected at the range corresponding to the first histogram minimum, i.e. 650 km., was of the order 10^7 m², yet the figure shows that such echoes did occur.

A study of the rate of growth curves obtained from ciné observations on burst echoes shows that, in general, the echo grows rapidly in amplitude to a maximum during an interval of about 2/25 sec. The subsequent decay process is much slower and is often closely exponential in form. Many echoes, on the other hand, do not show a smooth decay curve but one which oscillates in magnitude. Amplitude–time curves of this type have been obtained for a large number

of echoes originating from within the same small range interval. These curves show that the duration of a burst is not simply related to the maximum echoing area which it presents during its lifetime; they suggest that the processes of growth and decay of ionization must be subject to change from incident to incident, and that the manner of change in geometrical size of the ion cloud due to diffusion must also be variable.

§7. RELATIONSHIP BETWEEN APPARENT RATE OF OCCURRENCE OF BURSTS AND RADIO-FREQUENCY OF OBSERVING STATION

We have seen from figure 4 that the average level of activity during the eclipse experiments differed greatly from station to station, and it is of interest to examine whether this effect may be explained in terms of the known characteristics of the stations or whether the burst activity does indeed vary from locality to locality. In table 1 are set forth the average number of echoes per hour for each of the stations, together with the corresponding values of the integral $\int Q(R,h)dR$, for the appropriate range interval.

Table 1

Station	Frequency Mc/sec.	Echoes per hour between 120–220 miles$=a$	$\int Q(R,h)\,dR=b$	$m=a/b$
Class A				
Whalehead	25·10	802	109	7·4
Hillhead 1	22·69	607	80	7·6
Sango	25·10	256	33	7·7
Netherb'n	26·36	332	54	6·2
Class B				
Hillhead 2	42·27	236	109	2·2
Inverness	43·69	105	87	1·2
Tannach	42·27	88	94	0·9

The general conclusion drawn from this table is that the number of bursts observed by a station diminishes with increase of the radio frequency. We have found previously, however, that the echoing area distribution function expressed by equation (8) holds good for different radio-frequencies but with a different constant of proportionality. In order to interpret the table therefore it has appeared reasonable to modify equation (8) so as to include the effect of frequency explicitly and to write it in the form

$$n_\nu dA = A^{-3/2} cf(\nu)dA, \qquad \ldots\ldots(10)$$

where n_ν is the occurrence rate of bursts per unit area which present echoing areas between A and $A+dA$ to a radio wave of frequency ν, $f(\nu)$ is some function of the frequency and c is a constant independent of the frequency.

The occurrence rate of bursts which present echoing areas greater than some value A now becomes $A^{-\frac{1}{2}}cf(\nu)$, and (7) takes the form:

$$\text{Number of bursts in the range interval } R_1 \text{ to } R_2 \text{ observed to occur by a station operating on the radio-frequency } \nu: \propto f(\nu)\int_{R_1}^{R_2} Q(R,h)dR. \qquad \ldots\ldots(11)$$

Following this simple treatment the unknown function $f(\nu)$ is clearly proportional to the value of the ratio m given in the final column of table 1, and it

would follow that stations operating on approximately the same radio-frequency should yield the same m. Examination of table 1 shows that the m-values for the Class A stations are substantially of the same magnitude and that $f(\nu)$ for these stations will also be approximately constant since the frequency range is small. This correspondence suggests that the differences in station activities are not due to local variations of burst activity but may be adequately explained in terms of the physical factors of position and type of aerial, frequency and power radiated.

This question of local variation of burst activity was later examined in more detail by means of simultaneous observations made from two stations situated near Aberdeen and Weymouth respectively. It was not found that even this latitude separation produced any systematic activity differences.

The striking dependence of observed burst activity upon the radio-frequency employed is shown by comparing the results for the Class B stations with the figures for the A group. It is seen that doubling the radio-frequency produces a marked decrease in activity and a corresponding decrease in the m-values. The results are too rough to permit accurate deductions to be drawn and more work is clearly needed, but comparison of the average values of the two groups of stations show that $f(\nu)$ varies as $1/\nu^3$ and equation (10) becomes

$$n_s dA = \nu^{-3} A^{-3/2} c dA. \qquad \ldots\ldots(12)$$

It is clear that this apparent variation of burst activity with radio-frequency must be a consequence of the different echoing areas presented by a burst to radio waves of different frequencies and equation (12) leads immediately to the conclusion that the echoing area which a given burst presents to a wave of frequency ν varies inversely as ν^6. Using this result it may now be deduced from equation (1) that the amplitude of the echo signal which a given burst returns to a station, other things being equal, will vary as the fourth power of the wavelength of the radio waves employed. It is interesting to note that this result is in agreement with Eckersley's long wavelength observations on scatter clouds (1932).

§8. GENERAL CONCLUSIONS WITH REFERENCE TO THE METEORIC THEORY

Reference has already been made to the development of the theory associating the occurrence of transient bursts of ionization in the ionosphere with the incidence of meteors, and it is of interest to examine whether the present experimental results are in agreement with this theory.

In a comprehensive account of meteor phenomena given by Hoffmeister (1937) a distinction is drawn between the shower type of meteor which originates within the solar system and pursues an elliptic orbit about the sun and the so-called sporadic or cosmic meteor that arrives from outer space. It is now known that a large majority of the meteors incident upon the earth are of the sporadic type and that definite diurnal and seasonal cycles are associated with their rate of occurrence at a given point upon the earth. On a typical night the rate of occurrence of meteors increases to a maximum in the early hours of the morning and then the activity falls off until dawn. Similarly the nocturnal activity follows a smooth seasonal cycle with its maximum in the autumn and minimum in the spring.

These variations of optical meteor activity have been explained in terms of the change of position of the observing station relative to the vector which defines

the earth's instantaneous orbital velocity. This vector is defined by its point of intersection with the celestial sphere, a point which is termed the meteoric apex. The maximum number of observable meteor impacts will occur when the hemisphere containing the station is the leading face of the earth in its orbital motion and when the meteoric apex lies on the station's meridian. This occurs at 0600 hours; also at 1800 hours there will be a minimum of impacts since the station then lies on the rearward face of the earth and the meteoric apex is at its lowest point. Other configurations will yield activities between these extreme values. Similarly, the total daily activity will change with the elevation of the diurnal path of the meteoric apex over the celestial sphere, and thus show a seasonal variation with maximum in the autumn.

It will be immediately noticed that the description we have given of the diurnal and seasonal changes of ionospheric burst activity closely parallels the above account of the time variation of sporadic meteor activity. A maximum of activity in the early morning and a minimum about 1800 hours are features of the diurnal cycle required by the above theory of meteors and these same features have been found characteristic of the daily curve of burst activity (figure 1). The form of the seasonal curve of burst activity given in figure 3 shows similar agreement with the above description of the seasonal variation of meteor activity. The amplitude of the early morning peak also shows a seasonal change in conformity with the changing angle of elevation of the meteoric apex, and in figure 1 it is seen that the autumn and winter amplitudes exceed those for winter and summer.

The late morning peak of the burst cycle, however, cannot be explained in terms of known characteristics of meteors, so that some other approach is necessary. Appleton and Naismith (1947) have recently found a correlation between the noon level of burst activity and the state of ionization of the lower E region, and they suggest that the susceptibility of the ionosphere to impact ionization by a meteoric particle probably varies with the existing intensity of ionization. The ionized state of the E region is subject to solar control in a manner which is fully understood, so that, following Appleton and Naismith, it now appears reasonable to attribute the second maximum of the daily cycle of burst activity to the changing sensitivity of the lower E region. From this it would immediately follow that the amplitude of this peak should vary seasonally and be a maximum at the summer solstice; figure 1 shows that this is the case.

It thus appears that the diurnal and seasonal cycles of burst activity may be explained in terms of two main factors, (*a*) the rate of incidence of meteors upon the area of the ionosphere visible from the station, (*b*) sensitivity of the burst-producing layer to impact ionization. Effect (*a*) is measured by the orientation of the earth's orbital motion vector relative to the station, while effect (*b*) depends upon the state of ionization of the E region which is controlled by the elevation of the sun. Both these effects produce an afternoon fall in activity and so must result in the pronounced minimum which we have observed to be so consistently present on our daily curves.

A further point of correspondence with the meteor theory is that of the heights of occurrence of the bursts. Olivier (1925) and Hoffmeister (1937) have collected the results of a number of observers on the mean heights of appearance and disappearance of meteors; the former value varies between 160 and 90 km., and disappearance occurs below 120 km. and mostly at about

80 km. In particular, Trowbridge (1907) has carefully studied meteor trains and finds that they tend to occur at a mean height of 87 km. These various figures may be compared with our result that the bursts appear to produce their scattered radiation from within a thin layer of the ionosphere at a height of 86 km.

We therefore conclude that the evidence of the present investigation supports the theory that a large proportion of ionospheric bursts may be attributed to impact ionization produced by sporadic meteors. The absence of an eclipse effect, moreover, is not at variance with this view, for the eclipse of the solar radiation would cause only a temporary and slight reduction in the sensitivity of the critical layer at a time coincident with the normal fall in activity, so that the detection of such a slight change would be difficult by the burst-counting method. The daily variation in the rate of incidence of meteors would also obscure the small effect that might be expected.

ACKNOWLEDGMENTS

This work was carried out under R.A.F. auspices by the permission of Air Ministry and the A.O.C. of No. 60 Group; it was initially prompted by Sir Edward V. Appleton, F.R.S., to whom our thanks are due for helpful discussion and advice. We would also like to acknowledge the valuable assistance received from many R.A.F. personnel and particularly from F/O. N. R. Benzie.

REFERENCES

APPLETON, E. V., and NAISMITH, R., 1935, *Proc. Roy. Soc.* A, **150,** 685; 1940, *Proc. Phys. Soc.*, **52,** 402; 1946, *Nature, Lond.*, **158,** 936; 1947, *Proc. Phys. Soc.*, **59,** 461.

APPLETON, E. V., NAISMITH, R., and INGRAM, L. J., 1937, *Philos. Trans.* A, **236,** 191

APPLETON, E. V., and PIDDINGTON, J. H., 1938, *Proc. Roy. Soc.* A, **164,** 467.

ECKERSLEY, T. L., 1932, *J. Instn. Elect. Engrs.*, **71,** 405; 1937, *Nature, Lond.*, **140,** 846; 1940, *J. Instn. Elect. Engrs.*, **86,** 548.

ECKERSLEY, T. L., and FARMER, F. T., 1945, *Proc. Roy. Soc.* A, **184,** 196.

HEY, J. S., and STEWART, G. S., 1946, *Nature, Lond.*, **158,** 481.

HOFFMEISTER, C., 1937, *Die Meteore* (Berlin: Springer).

NAGAOKA, H., 1929, *Proc. Imp. Acad. Tokyo*, **6,** 5.

OLIVIER, C. P., 1925, *Meteors* (Baltimore: Williams and Wilkins).

PIERCE, J. A., 1938, *Proc. Inst. Radio Engrs.*, **26,** 892.

SCHAFER, J. P., and GOODALL, W. M., 1932, *Proc. Inst. Radio Engrs.*, **20,** 1941.

SKELLETT, A. M., 1932, *Proc. Inst. Radio Engrs.*, **20,** 1933; 1935, *Ibid.*, **23,** 132; 1938, *Nature, Lond.*, **141,** 472.

TROWBRIDGE, C. C., 1907, *Astrophys. J.*, **26,** 95.

Measurements of the Interaction of Radio Waves in the Ionosphere

BY L. G. H. HUXLEY, H. G. FOSTER AND C. C. NEWTON

Department of Electrical Engineering, University of Birmingham

MS. received 28 *September* 1947

ABSTRACT. This paper describes measurements of radio wave interaction, in special test transmissions, between the B.B.C. transmitters at Droitwich and Lisnagarvey (N. Ireland). It is shown that the theoretical formula $M=M_0/[1+(n/G\nu)^2]^{\frac{1}{2}}$ accurately describes the dependence of the interaction modulation on the modulation frequency ($n/2\pi$). The electronic collisional frequency ν at the seat of interaction is deduced from $G\nu$ and the laboratory value of G. Investigations of the phase of the modulation are also described.

The paper contains a sketch of the theory of wave interaction. There is also an addendum dealing with the possible use of wave interaction as a tool in ionospheric research.

§1. INTRODUCTION

THIS paper describes the results of a number of observations of the phenomenon of non-resonant radio wave interaction in the ionosphere in special test transmissions. The equipment employed was similar to, but an improvement on, that used by us in earlier observations on gyro-interaction that have already been reported (Huxley, Foster and Newton 1947). The test transmissions which concern us in this paper were arranged by Mr. J. A. Ratcliffe of the Cavendish Laboratory in consultation with the Engineering Research Department of the British Broadcating Corporation, and through their courtesy we were informed of these transmissions. Mr. Ratcliffe also informed us of his method for roughly locating the region of interaction described in §5 below.

A special effort was made in our observations to obtain accurate measurements of the depth of interaction modulation as a function of audio-frequency.

The transmitters used in the tests were Droitwich (170 kw., 200 kc/s., 80% modulation) and Lisnagarvey, N. Ireland (60 kw., 1050 kc/s., plain carrier), although Droitwich was replaced by Ottringham, near Hull (170 kw., 167 kc/s.), on a few occasions. The point of observation was in Selly Oak, Birmingham. Their relative positions are shown in figure 1.

Before proceeding to a description of the experimental results, a brief account of the phenomenon of wave interaction and the elements of the theory are given in order that the significance of the observations may be the better appreciated.

§2. DESCRIPTION OF THE PHENOMENON AND NOMENCLATURE

Consider two radio transmitters; the one, which we shall call the interfering transmitter, radiates at high power a modulated carrier, whereas the carrier of the other, called the wanted transmitter, is unmodulated. An observer, so placed that he can receive the components of the transmission from the wanted station that have passed through the E region of the ionosphere in the vicinity of the interfering station, finds that, in general, the modulation of the interfering transmission is impressed upon the wanted transmission in the course of its passage through the E region. This phenomenon is called Wave Interaction (previously the Luxembourg Effect). The phenomenon was first reported by Tellegen (1933) and the first satisfactory attempt at a theoretical explanation was made by Bailey and Martyn (1934). Later Bailey (1937) revised the original theory and gave a more rigorous and complete treatment. Systematic investigations of the phenomenon were made by Grosskopf (1938).

§3. SIMPLIFIED THEORY OF WAVE INTERACTION

The basic ideas of the theory given by Bailey and Martyn and later elaborated by Bailey are the following. The absorption coefficient k_0 of the ionosphere for the wanted wave whose frequency is $p/2\pi$ at a place where the collisional frequency is ν is known to be (neglecting the earth's magnetic field)

$$k_0 = \frac{2\pi \boldsymbol{e}^2 \boldsymbol{c} N}{\boldsymbol{m}} \left(\frac{\nu}{\nu^2 + p^2} \right) \text{cm}^{-1}, \qquad \ldots\ldots(1)$$

where $\boldsymbol{e}$ is the electronic charge in E.M.U., $\boldsymbol{m}$ the mass of the electron, $\boldsymbol{c} = 3 \times 10^{10}$ and N is the electronic concentration in cm^{-3}. The absorption coefficient is thus a function of the electronic collisional frequency and the frequency of the wanted wave.

The E region in the vicinity of the interfering station (frequency $p_1/2\pi$) is irradiated by the electric field of the latter,

$$E_1(1+D\cos nt)\cos p_1 t, \qquad \ldots\ldots(2)$$

where D is the modulation depth and $n/2\pi$ the modulation frequency. Suppose first that the modulation D is zero. The field $E_1\cos p_1 t$ then supplies power W to an electron at a specified region in the ionosphere. When W is zero, the kinetic energy of agitation of an electron is the same as the thermal energy of agitation Q_t of the molecules of the gas. When, however, power W is supplied to an electron at a steady rate, a new condition of dynamical equilibrium results in which the mean kinetic energy $\overline{Q}$ of the electron exceeds that of the gas molecules, so that we may write $\overline{Q}=k_T Q_t$, where k_T is Townsend's energy factor.

In this condition, let the collisional frequency of the electron be $\bar{\nu}$ and $\gamma\overline{Q}$ the mean energy lost per collision with gas molecules. Then

$$\gamma\bar{\nu}\overline{Q}=W. \qquad \ldots\ldots(3)$$

When the collisions between electrons and gas molecules resemble those between smooth rigid spheres, it is known that with a Maxwellian distribution of velocities

$$\gamma=2{\cdot}66(\boldsymbol{m}/M')(1-1/k_T)=2{\cdot}66(\boldsymbol{m}/M')(1-\nu_t^2/\bar{\nu}^2),$$

where M' is the mass of a molecule of the gas and ν_t is the collisional frequency of the electron when in thermal equilibrium with the gas ($W=0$).

Bailey has supposed that a similar formula is applicable to an electronic motion in a diatomic gas when $2{\cdot}66\boldsymbol{m}/M'$ is replaced by a factor G which is characteristic of the gas. Thus,

$$\gamma=G(1-1/k_T)=G(1-\nu_t^2/\bar{\nu}^2). \qquad \ldots\ldots(4)$$

In general, G greatly exceeds $2{\cdot}66\boldsymbol{m}/M'$, as for instance in air, where $2{\cdot}66\boldsymbol{m}/M'=5\times10^{-5}$ but $G\simeq10^{-3}$.

When the coefficient of modulation D in equation (2) is zero, suppose E_1 to supply power to an electron at the mean rate

$$W=AE_1^2, \qquad \ldots\ldots(5)$$

where A depends on the frequency $p_1/2\pi$, the collisional frequency $\bar{\nu}$ and $\boldsymbol{e}^2/\boldsymbol{m}$ (Appendix).

When D is not zero the power supplied to an electron becomes time-dependent as follows:

$$W(t)=W(1+D\cos nt)^2=\overline{W}+2W(D\cos nt+\tfrac{1}{4}D^2\cos 2nt), \qquad \ldots\ldots(6)$$

where

$$\overline{W}=(1+D^2/2)W. \qquad \ldots\ldots(7$$

The energy of agitation Q now becomes a function of the time which we determine.

Let $\overline{Q}$ be the mean value of Q when power $\overline{W}$ is communicated to each electron, and Q_t its value when $\overline{W}=0$ (thermal equilibrium). It follows from (4) and (6) with $k_T=\overline{Q}/Q_t=(\bar{\nu}/\nu_t)^2$, that

$$dQ/dt+(G\bar{\nu})(Q-Q_t)=W(t). \qquad \ldots\ldots(8)$$

Similarly,

$$(G\bar{\nu})(\overline{Q}-Q_t)=\overline{W}=W(1+D^2/2). \qquad \ldots\ldots(9)$$

Hence

$$dQ/dt+(G\bar{\nu})(Q-\overline{Q})=2W[D\cos nt+\tfrac{1}{4}D^2\cos 2nt], \qquad \ldots\ldots(10)$$

whence

$$Q=\overline{Q}\left[1+\frac{2W}{\overline{Q}(G\bar{\nu})}\left\{\frac{D\cos(nt-\phi)}{(1+(n/G\bar{\nu})^2)^{\frac{1}{2}}}+\frac{D^2\cos(2nt-\phi')}{4(1+(2n/G\bar{\nu})^2)^{\frac{1}{2}}}\right\}\right], \qquad \ldots\ldots(11)$$

where

$$\tan\phi=(n/G\bar{\nu}); \qquad \tan\phi'=(2n/G\bar{\nu}). \qquad \ldots\ldots(12)$$

Thus the agitational energy Q fluctuates about its mean value $\bar{Q}$ at a frequency $(n/2\pi)$. Since the collisional frequencies ν and $\bar{\nu}$ corresponding to Q and $\bar{Q}$ satisfy the relation $(\nu/\bar{\nu})^2=(Q/\bar{Q})$, and since in practice the coefficient $2W/[\bar{Q}(G\bar{\nu})]\ll 1$ in (11), it follows that ν is modulated as follows:

$$\nu=\bar{\nu}\left[1+\frac{W}{\bar{Q}(G\bar{\nu})}\left\{\frac{D\cos(nt-\phi)}{(1+(n/G\bar{\nu})^2)^{\frac{1}{2}}}+\frac{D^2\cos(2nt-\phi')}{4(1+(2n/G\bar{\nu})^2)^{\frac{1}{2}}}\right\}\right]. \quad \ldots\ldots(13)$$

Thus the ratio of the amplitude of the second harmonic to that of the fundamental in the modulation of Q and ν is

$$\tfrac{1}{4}D\{n^2+(G\bar{\nu})^2\}/[4n^2+(G\bar{\nu})^2]. \quad \ldots\ldots(14)$$

It is the custom to neglect the second harmonic but its existence in the correct proportion required by (14) has been established in measurements of wave interaction (Huxley, Foster and Newton 1947).

In what follows we consider the fundamental term alone and write (13) $\nu=\bar{\nu}+\nu_n$ with

$$\nu_n=\frac{\bar{\nu}WD}{\bar{Q}(G\bar{\nu})}\frac{\cos(nt-\phi)}{[1+(n/G\bar{\nu})^2]^{\frac{1}{2}}}. \quad \ldots\ldots(15)$$

It follows from (9) and (14) that the coefficient $WD/Q(G\bar{\nu})$ in (13) may be written in the alternative forms

$$\frac{WD}{\bar{Q}(G\bar{\nu})}=\frac{WD}{\bar{W}+Q_{\mathrm{i}}(G\bar{\nu})}=\frac{D}{[(1+D^2/2)+Q_{\mathrm{i}}(G\bar{\nu})/\bar{W}]}, \quad \ldots\ldots(16)$$

and

$$WD/\{\bar{Q}(G\bar{\nu})\}=D(k_{\mathrm{T}}-1)/(1+D^2/2)k_{\mathrm{T}}. \quad \ldots\ldots(17)$$

When the wanted wave traverses a region in which the collisional frequency is modulated as in (15) by the field of the interfering station, the absorption coefficient (equation (1)) also becomes modulated as follows:

$$k=k_0+\nu_n\,dk_0/d\nu. \quad \ldots\ldots(18)$$

Suppose the modulation of k to occur within a region which is traversed by the wanted wave along a ray of length s. The emergent amplitude of the wanted wave in terms of its amplitude on entering the region is

$$E=E_0\exp\left(-\int_0^s k\,ds\right)=E_0\exp\left(-\int_0^s k_0\,ds\right).\exp\left(-\int_0^s \nu_n(dk_0/d\nu)\,ds\right), \quad \ldots\ldots(19)$$

Since the total attenuation of the wave is in practice not excessive, the exponent $-\int_0^s k_0\,ds$ is of the order of magnitude of unity or less. The increment in the absorption is still less and we may write $\int_0^s \nu_n(dk_0/d\nu)\,ds\ll 1$; consequently (19) may be written,

$$E=E_0\exp\left(-\int_0^s k_0\,ds\right)\left[1-\int_0^s \nu_n(dk_0/d\nu)\,ds\right]. \quad \ldots\ldots(20)$$

Thus, the depth of the impressed modulation, expressed as a fraction of the received amplitude of the wanted wave is

$$M=-\int_0^s \nu_n(dk_0/d\nu)\,ds. \quad \ldots\ldots(21)$$

Suppose for simplicity that k_0 and $\bar{\nu}$ are constant along the path of integration through the region of interaction. This implies either that the path of integration lies at the top of the trajectory of the ray, or that, if the interaction region lies on the incident or reflected ray, it is effectively localized. The expression for M then simplifies (from (16)) to

$$M = \frac{D\bar{\nu}(dk_0/d\nu)s}{[1+D^2/2+(G\bar{\nu})Q_t/W]} \frac{\cos(nt-\phi)}{[1+(n/G\bar{\nu})^2]^{\frac{1}{2}}}. \qquad \ldots\ldots(22)$$

Which alternative is the more appropriate is to be decided from experience. In any case, it is supposed that where interaction occurs $p^2 \gg \bar{\nu}^2$, so that according to (1) the absorption coefficient is proportional to $\bar{\nu}$. The total absorption of the wanted wave in nepers is $\int_0^s k_0\,ds$, which is, therefore, of the form $C\bar{\nu}$ where C is determined by medium and the path. The term $s(dk_0/d\nu_0)\bar{\nu}$ in (22) may be replaced by $(C\bar{\nu}s)$, which can be estimated from the reflection coefficient of the E region for the wanted wave.

The expression (22) for M shows that in this case the dependence of M on the modulation frequency is relatively simple and of the form

$$M = M_0/[1+(n/G\bar{\nu})^2]^{\frac{1}{2}}, \qquad \ldots\ldots(23)$$

where, from (22), $M_0 = DCs\bar{\nu}/[(1+D^2/2)+Q_t(G\bar{\nu}/W])$.

Further, the interaction modulation lags in phase with respect to the interfering field E_1 at the region of modulation by the angle

$$\phi = \tan^{-1}(n/G\nu). \qquad \ldots\ldots(24)$$

In the measurements described below special attention was paid to the investigation of M as a function of n and it was found that formula (23) describes accurately what is observed when $n/2\pi$ does not exceed 750 c/s. From such measurements the quantity $G\bar{\nu}$ may be deduced in each instance.

In practice, the term $(G\bar{\nu})Q_t/W \gg (1+D^2/2)$ in (22) so that M_0 is effectively equal to $WDCs\bar{\nu}/Q_t(G\bar{\nu})$.

§4. MEASUREMENT OF INTERACTION MODULATION AS A FUNCTION OF MODULATION FREQUENCY

(i) *Calibration of equipment*

Because the depths of interaction modulation are relatively small in practice, it was necessary to use measuring equipment capable of giving reliable results down to modulation depths of about 0·1% on a carrier received from a relatively distant (300–400 km.) medium wave station. It was decided therefore to use not an oscilloscope but a specially calibrated receiver and to obtain the carrier voltage and the audio-frequency voltage at the detector from meter readings, from which the modulation depth could be obtained from calibration curves. The receiver was a modified commercial communications receiver which covered the required range of wavelengths. The detector stage was modified to feed directly into a valve voltmeter circuit which was built into the receiver chassis, and which was calibrated to provide the carrier level at the detector in the form of a meter reading.

The audio frequency voltages from the detector were fed, through suitable filter circuits, to a cathode follower, the output of which was connected directly to a Marconi wave analyser which was available in the laboratory. The advantage of a wave analyser lies in its narrow bandwidth of only 4 c/s. which enables the

wanted modulation tone from the detector to be measured relatively free from background noise and interference.

Through a laboratory calibration it was possible to use the direct readings of the wave analyser in combination with those of the carrier-level meter mentioned above to obtain the required modulation depth immediately. The wave analyser had, in fact, been originally used to measure the second harmonic of the interaction modulation shown in equation (13) as mentioned in the paragraph that follows it.

The calibration of the system to read modulation depths direct was achieved by applying to the input, a radio frequency carrier whose frequency was near that of the wanted transmission. This carrier was modulated by a series of pure tones covering the range of frequencies of 50 to 2000 c/s. The modulation depths, for the purposes of calibration were set in turn at $2\frac{1}{2}\%$, 5% and 10%. Later this calibration was found to agree with a calibration tone of known depth specially imposed upon the carrier of the wanted transmission radiated by Lisnagarvey (and Westerglen on another occasion).

The laboratory calibration unit comprised a crystal oscillator driving a power amplifier which was modulated by the appropriate tone.

By extrapolation of the calibration curves which are straight lines, modulation depths down to 0·1% could be confidently assessed.

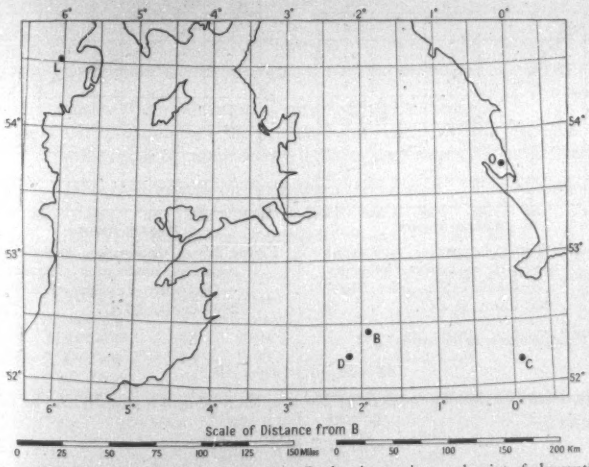


Figure 1. Gnomonic projection centred on B, showing stations and point of observation.

(ii) *Experimental results*

The transmitters concerned are shown in figure 1, in which D, L and O represent respectively Droitwich, Lisnagarvey and Ottringham. The point of observation was in Birmingham and is indicated by B. The distance LB is 340 km., whereas DB is 20 km. and the angle LBD is approximately a right angle.

The transmitter at D radiated 170 kw. in the form of a carrier on 200 kc/s. modulated by a series of pure tones to a depth of 80%, while at the same time the transmitter L radiated 60 kw. as a plain carrier on a frequency of 1050 kc/s.

The times of observation were between the hours of 2230–0300 G.M.T. on the nights of June 15/16 and 16/17, July 3/4, 4/5, 5/6 and 6/7, and August 6/7, 7/8 and 8/9, 1947.

At intervals during the night, pulse transmissions were sent out by Lisnagarvey which showed that the ground wave received in Birmingham was relatively weak, and that the predominant wave received during the early night was a single reflection from the E region, often accompanied later by other forms of reflected waves from the E or F regions with relative strengths varying continually.

In the test runs, in which M was determined as a function of $(n/2\pi)$ the modulation frequency, Droitwich was modulated in turn, for two minutes each, at the frequencies 50, 75, 100, 150, 200, 300, 400, 500, 750, 1000, 1250, 1750 and 2000 c/s.

The modulation depths were found from simultaneous readings of the carrier level meter and the wave analyser as described in the previous section, using A.V.C. to hold the carrier level constant at the detector. It was confirmed that the use of A.V.C. did not affect the measured modulation depth, except to improve the accuracy.

Two examples of the results of such measurements are shown in the curves of figures 2 (*a*) and (*b*). They refer to the times stated on the figures.

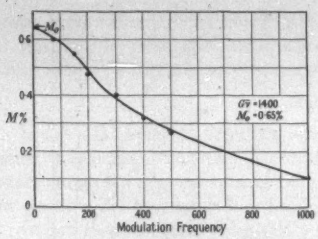


Figure 2 (*a*). Interaction modulation depth *vs* modulation frequency —Droitwich—Lisnagarvey 0230–0300 G.M.T. 16/6/47.

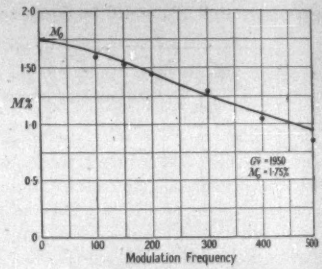


Figure 2 (*b*). Interaction modulation depth *vs* modulation frequency —Droitwich—Lisnagarvey 2330–2400 G.M.T. 3/7/47.

If the theoretical formula (23),

$$M = M_0/[1 + (n/G\bar{\nu})^2]^{\frac{1}{2}} \qquad \ldots\ldots(25)$$

actually describes the dependence of M on n, then M and n should also be related as follows:—

$$n^2 = M_0^2(G\bar{\nu}/M)^2 - (G\bar{\nu})^2, \qquad \ldots\ldots(26)$$

in which M_0 is the value of M as $n \to 0$.

Thus, the curve which represents n^2 as a function of $1/M^2$ is a straight line whose intercept on the n^2 axis is $(G\bar{\nu})^2$. In figures 3 (*a*) and (*b*) the experimental data exhibited in figures 2 (*a*) and (*b*) are plotted with n^2 as a function of $1/M^2$. It can be seen that for frequencies $n/2\pi$ from 50 to 500 c/s. a linear relation subsists; consequently, at the lower frequencies $n/2\pi$, where the measurements are more accurate, equation (25), does in fact closely describe the experimental facts.

Evidently the quantity $G\bar{\nu}$ may be found from the intercept on the n^2 axis of the straight line through the experimental points. Further, M_0 is found from the intercept on the $1/M^2$ axis.

It is therefore necessary only to give the zero frequency modulation M_0 and the value of $G\bar{\nu}$ in order to specify completely the modulation M at any frequency $n/2\pi$.

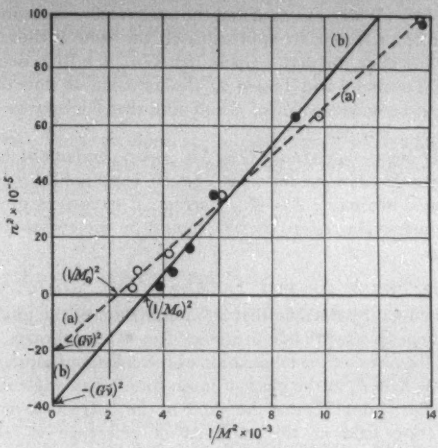


Figure 3. Points in figure 2 replotted to exhibit $1/M^2$ as a function of $n^2=(3\pi f)^2$.

In what follows the modulation is specified in this manner.

Tables 1 and 2 summarize the experimental results on the dependence of M on n. The values of $G\bar{\nu}$ ranged from 1100 to more than 2000 in these tests.

Table 1. Observed depths on interaction modulation. Wanted station—Lisnagarvey (1050 kc/s.)

Interfering transmitter—Droitwich (200 kc/s.; 170 kw.) 80%.

Date	Time G.M.T.	M_0 (%)	$G\bar{\nu}$	Date	Time G.M.T.	M_0 (%)	$G\bar{\nu}$
June 16	0230–0300	0·6	1400	July 5	2230–2300	1·63	2320
,, 16	2300–2330	1·75	1480	,, 5	2330–2400	2·5	—
,, 17	0230–0300	0·9	irregular	,, 6	2230–2300	1·2	2860
July 3	2330–2400	1·75	1950	,, 6	2330–2400	1·4	1480
,, 4	2230–2300	1·75	1870	,, 7	0130–0200	2·1	1480
,, 4	2330–2400	1·85	1520				

Interfering transmitter—Ottringham (167 kc/s.; 200 kw.) 80%.

Date	Time G.M.T.	M_0 (%)	$G\bar{\nu}$
June 16	0200–0230	0·58	1550
,, 17	0200–0230	0·45	1480

Table 2. Observed depths of interaction modulation. Wanted station—Lisnagarvey (1050 kc/s.)

Interfering transmitter—Droitwich (200 kc/s.; 170 kw.).

Date	Time G.M.T.	M_0 (%)	$G\bar{\nu}$	h (km.)	Date	Time G.M.T.	M_0 (%)	$G\bar{\nu}$	h (km.)
Aug. 6	2300–2328	2·6	1550	75	Aug. 7	2300–2328	2·6	1340	95
,, 6	2340–2348	2·6	1100	68	,, 7	2340–2348	2·1	—	86
,, 7	0010–0018	1·55	1050	70	,, 8	0010–0018	1·9	1720	86
,, 7	0040–0048	2·2	1570	88	,, 8	0040–0048	2·25	1870	86
,, 7	0130–0158	1·9	1450	88	,, 8	0130–0158	2·1	1220	97
,, 7	0210–0218	—	—	83	,, 8	0210–0218	1·5	1790	86
,, 7	0240–0248	2·0	1950	83	,, 8	0240–0248	1·5	—	86
					,, 9	0100–0128	0·77	1140	82

h=height of interaction.

It was found in an isolated observation of wave interaction between Westerglen (767 kc/s.) near Falkirk and Ottringham (167 kc/s.–170 kw.) on 8th May (0030–0100) that $M_0 = 4\%$ and $G\bar{\nu} = 1800$. Thus, $G\bar{\nu}$ is of the order, in these experiments, of $1{\cdot}5 \times 10^3$. If it be assumed that the temperature and composition of the air at the region of interaction are approximately the same as those at the ground, then, as Bailey has done, we may suppose that $G = 2{\cdot}6 \times 10^{-3}$, deduced from the experiments of Townsend and Tizard on the diffusion of slow electrons in air. Then from the experimental value of $G\bar{\nu}$ we find that $\bar{\nu}$ is approximately 6×10^5, and exceeds ν_0 by a few per cent.

Recent experiments to determine G by Mr. Zaazou and one of us in this department have shown that its value for air is $1{\cdot}3 \times 10^{-3}$ and not $2{\cdot}6 \times 10^{-3}$ as supposed hitherto. Thus $\bar{\nu}$ becomes $1{\cdot}2 \times 10^6$. Further, if oxygen is present as atomic oxygen at the interaction region, then G should be somewhat smaller than the laboratory value.

§ 5. THE PHASE OF THE INTERACTION MODULATION

It was pointed out by Ratcliffe that an examination of the phase of the interaction modulation should provide a further test of the theory. According to equations (12), the phase of the modulation of the collision frequency lags on that of the interfering field E_1 at the place of interaction by an angle $\phi = \tan^{-1}(n/G\bar{\nu})$. It follows therefore, from (22) that the phase of the interaction modulation, with that of interfering field as the standard of reference as above, lags by $\phi = \tan^{-1}(n/G\nu_0)$ or $\tan^{-1}(n/G\nu_0) \pm 180°$ according as $(dk_0/d\nu)$ in (21) has a negative or a positive sign. Since, from (1), k_0 is a maximum when $p = \nu$, and $(dk/d\nu)$ is negative when $\nu > p$, but positive when $\nu < p$, it follows that the very low frequency modulation $M_0(n \to 0)$ should vibrate in phase with the interfering modulation if interaction occurs below the height where $\nu = p$, but should vibrate in anti-phase if injected in a region at a greater height.

It was found that the zero frequency modulation was anti-phased with respect to that of the interfering field. It may therefore be concluded that interaction takes place above the level at which $\nu = p$ (where $p/2\pi$ is the frequency of the wanted wave).

In figure 4, D represents the interfering transmitter (Droitwich), B the point of observation (Selly Oak, Birmingham) and I the centre of a region where interaction occurs. Thus, IB is a ray from the wanted station (Lisnagarvey) that reaches B. Let DB $= l_1$ (20 km.) and put $l_2 =$ DI + IB. Take as reference phase for the modulation tones that of the interfering modulation at D, its source. The phase lag ψ_1 of this modulation when received at B on the carrier of the interfering station is

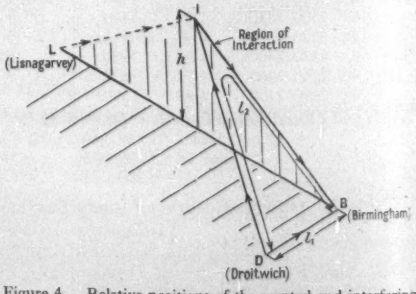


Figure 4. Relative positions of the wanted and interfering stations and of the reception point.

$$\psi_1 = 360 f l_1 / v \text{ degrees,} \qquad \ldots\ldots(27)$$

where the modulation frequency $f = n/2\pi$ c/s. and $v = 3 \times 10^5$ km/sec.

Similarly, the phase of the interaction modulation generated at I, when received at B on the carrier of the wanted wave, lags on the reference phase at D by

$$\psi_2 = 360fl_2/v + \tan^{-1}(2\pi f/G\bar{v}) \pm 180 \text{ degrees.} \quad \ldots\ldots(28)$$

Thus, if the audio outputs of identical receivers are compared on a double-beam oscilloscope, the relative phase lag at B of the interaction modulation with respect to the interfering modulation, which is

$$(\psi_2 - \psi_1) = 360f(l_2 - l_1)/v + \tan^{-1}(2\pi f/G\bar{v}) \pm 180 \text{ degrees,} \quad \ldots\ldots(29)$$

may be measured. If the receivers are not identical the phase shift introduced by each at any frequency f is easily determined and the difference used to correct the value of $(\psi_2 - \psi_1)$ given by the oscilloscope.

Let ψ be the frequency-dependent part of (29), that is

$$\psi = 360f(l_2 - l_1)/v + \tan^{-1}(2\pi f/G\bar{v}_0). \quad \ldots\ldots(30)$$

It follows that $\psi = 0$ when $f = 0$, and that

$$\psi \to 360f(l_2 - l_1)/v + 90°, \text{ as } f \to \infty.$$

Thus the curve representing ψ as a function of f passes through the origin and possesses an asymptote whose slope is

$$f \to \infty (d\psi/df) = 360(l_2 - l_1)/v, \quad \ldots\ldots(31)$$

and whose intercept with the ψ-axis is 90°.

When the measured values of ψ are plotted against f, the distance l_2 may be obtained from the slope of the asymptote as follows:

$$l_2 = l_1 + \tfrac{1}{12}(d\psi/df)10^4. \quad \ldots\ldots(32)$$

It follows that if the elevation of the downcoming ray IB and the angle LBD are known, then the point I may be determined. This method of fixing I from (30) and (32) was pointed out to us by J. A. Ratcliffe.

Figure 5 shows that a typical experimental curve representing ψ as a function of f conforms closely to what is to be expected from (30). On the assumption that the region of interaction I lies on a ray reflected at a height of 100 km., it was found that in most cases the apparent interaction point obtained from (32) was located at a height of (86 ± 5) km. The method therefore provides a rough estimate only of the height.

The most stable readings of amplitude and phase are obtained when one of the waves reflected from the

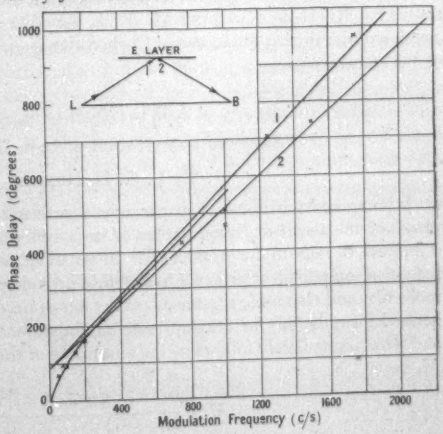


Figure 5.

Note. Inset: Point 1 obtained from asymptote 1.
Point 2 obtained from asymptote 2.

ionosphere is predominant. In these tests the predominant wave was the wave undergoing a single reflection at the E region as was shown by pulse transmissions from Lisnagarvey. However, except during the early evening, other waves reflected from the F or E regions appeared in varying strengths, and the interaction modulation on these components was able, on occasions, to shift the phase in an irregular fashion and to alter the amplitude of the interaction as measured at the point of observation.

It was noted that when A.V.C. was employed in measurements of modulation depth, the audio-frequency amplitude showed sudden departures from a steady value, although the carrier was held fixed. The steady level was assumed to correspond to the single E wave and the fluctuations to the presence of other sky waves.

§6. REMARK ON THE MAGNITUDE OF THE INTERACTION

The maximum depths of interaction modulation M_0 measured in these tests were small, as can be seen from tables 1 and 2, and we believe that this relatively small value is to be attributed to the relatively high frequency of the wanted station (1050 kc/s.).

It is of interest to compare the values of M_0 listed in tables 1 and 2 with those obtained in other transmissions.

Thus, van der Pol and van der Mark found that the interaction M_0 impressed on Beromunster (556 kc/s.) by Luxembourg (150 kw.) was about 8%, and as mentioned below table 1, we observed that Ottringham (100 kw.) impressed upon Westerglen (767 kc/s.) a modulation M_0 of about 4%.

In these tests the modulations M_0 impressed upon Lisnagarvey (1050 kc/s.) by Droitwich (170 kw.) were less than 2%. Thus, M_0 appears to diminish as p increases.

ACKNOWLEDGMENT

The authors wish to express their appreciation of the co-operation and interest shown by Mr. H. L. Kirke and Mr. R. A. Rowden of the Engineering Research Section of the British Broadcasting Corporation in providing them with schedules of the transmissions.

ADDENDUM

Suggested measurements of wave interaction

By L. G. H. HUXLEY

It is explained, in the above paper, how a determination of the amplitude and phase of the impressed modulation M as a function of modulation frequency $n/2\pi$ leads to value of the product $G\bar{v}$. It is natural to enquire whether the phenomenon of interaction of radio waves is able to provide estimates of the value of other molecular and electronic parameters of the ionosphere. To this end, consider the general formula (22) for the impressed modulation. According to this formula, and what immediately follows it, the amplitude of the zero frequency modulation M_0 is

$$M_0 = \frac{DC\bar{\nu}s}{(1+D^2/2+(G\bar{\nu})Q_t/W)}. \qquad \ldots\ldots(A\,1)$$

The mean carrier power W supplied by the interfering field E_1 to an electron is proportional to the radiated power P of the interfering transmitter, so that in

general M_0 is not proportional to the radiated power P, but becomes so in effect when the term $(G\bar{\nu})Q_t/W \gg 1$.

We therefore first consider the order of magnitude of $(G\bar{\nu})Q_t/W$ in the practical case where $P=200$ kw. and the frequency $p_1/2\pi = 2\times10^5$. We may consider, with good approximation to the truth, that the aerial with its image radiates as a Hertz doublet.

The amplitude E_1 of the field produced at a point at height h in the E layer with an angle of elevation α with respect to the transmitter is given by

$$E_1^2 = 4{\cdot}5\times10^{-6}(P\sin 2\alpha)/h^2 \text{ (volts/cm)}^2, \qquad \ldots\ldots(A\,2)$$

where P is in kw., h in km. and E_1 in v/cm.

It follows that the amplitude (peak) of the field E_1 when $h=100$ km., and $P=200$ kw. is $2{\cdot}12\sqrt{2}\times10^{-4}\sin 2\alpha$ v/cm.

To estimate W, assume a collisional frequency $\bar{\nu}=10^6$. We may suppose therefore that $p_1^2=(4\pi\times10^5)^2 \simeq \bar{\nu}^2$, and proceed to estimate $W=AE_1^2$, with

$$A \simeq \{\tfrac{1}{2}\mathbf{e}^2\bar{\nu}/\mathbf{m}(p_1^2+\bar{\nu}^2)\}\cos^2\theta$$

We find $W = 1{\cdot}8\times10^{-11}\cos^2\theta\sin 2\alpha$ erg/sec. If the temperature of the region where interaction occurs lies between 200° and 300° K., then Q_t is approximately 5×10^{-14} erg. Since $G\bar{\nu}\simeq 1{\cdot}5\times10^3$, we find

$$(G\bar{\nu})Q_t/W \simeq 4{\cdot}2/\cos^2\theta\sin 2\alpha.$$

The maximum modulation is obtained when the point of observation is so chosen, if possible, that θ, the angle between E_1 and the earth's magnetic field, is zero and α is 45°. The term $(G\bar{\nu})Q_t/W$ in (A 1) is then approximately $(1+D^2/2)$ and the impressed modulation M_0 would be less at $P=200$ kw. than would be predicted from its value when P is small, on the assumption that M_0 is proportional to P.

If, however, the maximum power P were notably less than 200 kw. or the angles θ and α appreciably different from zero and 45° respectively, then $(G\bar{\nu})Q_t/W$ at maximum power would certainly exceed $(1+D^2/2)$ by a factor of eight or more. An observer, unless his measurements of modulation M_0 were very accurate, would judge M_0 to be proportional to P up to the maximum value of P.

If, on the other hand, a very powerful transmitter capable of radiating 300 kw. or more, or a strong gyro-transmitter, were employed to produce the interfering field, then measurements of M_0 by the sensitive method described in §4.1 should reveal the lack of proportionality between M_0 and P under optimum geometrical conditions.

Such measurements should also provide additional information about the physical properties of the E layer, as may be appreciated from the following considerations.

For equation (A 2) write $E_1^2=KP$, then, $W=AE_1^2=AKP=bP$, where

$$b = 10^{-6}\frac{\mathbf{e}^2}{\mathbf{m}}\cdot\frac{\bar{\nu}}{(p_1^2+\bar{\nu}^2)}\cdot\frac{\cos^2\theta\sin 2\alpha}{h^2}. \qquad \ldots\ldots(A\,3)$$

Write (A 1) in the form

$$M_0 = a/[1+D^2/2+(G\bar{\nu})Q_t/bP], \qquad \ldots\ldots(A\,4)$$

where, from (1),

$$a = DC\bar{\nu}s. \qquad \ldots\ldots(A\,5)$$

Expression (A 4) is equivalent to

$$1/M_0 = (1+D^2/2)/a + (G\bar{\nu})Q_t/abP. \qquad \ldots\ldots(A\,6)$$

If, therefore, the experimental values of $1/M_0$ were plotted against $1/P$, the points should fall near a line whose slope is $(G\bar{\nu})Q_v/ab$ and whose intercept on the $1/M_0$ axis is $(1+D^2/2)/a$.

Thus a and $(G\bar{\nu})Q_v/b$ could be found individually. Since $G\bar{\nu}$ is obtained from measurements of M_0 as a function of modulation frequency $n/2\pi$, it follows that we obtain, from (A 6), a and Q_v/b, and the absorption coefficient $C\bar{s\nu}$ from a. If, however, proportionality between M_0 and P persists at very large values of P, then it is possible to set lower limits only to $C\bar{s\nu}$, the total absorption in nepers along the path s.

Appendix

An important quantity that appears in the theory of wave interaction as outlined in §3 is W, the mean power given to a free electron in a gas by a high frequency alternating electric field $E_1\cos p_1 t$, which for simplicity was written in equation (5) as $W=AE_1^2$. Expressions for W were derived by Bailey (1937), and independently at about the same time, slightly more accurate expressions were found by one of us (Huxley 1937). (Note that Bailey's w is $1/\nu$ times the W used in §3. It is the work communicated per collision.) It is found that the high frequency conductivity of a gas containing N electrons per cm^3 is

$$\sigma_{\parallel}=\frac{2}{3}\frac{Ne^2}{m}\left(\frac{\nu}{p_1^2+\nu^2}\right)\left(1+\frac{p_1^2}{p_1^2+\nu^2}\right) \text{ E.M.U.} \qquad \ldots\ldots(A\,7)$$

and

$$\sigma_{\perp}=\frac{1}{3}\frac{Ne^2}{m}\nu\left\{\frac{1}{\nu^2+(p_1-\omega)^2}\left[1+\frac{(p_1-\omega)^2}{\nu^2+(p_1-\omega)^2}\right]\right.$$
$$\left.+\frac{1}{\nu^2+(p_1+\omega)^2}\left[1+\frac{(p_1+\omega)^2}{\nu^2+(p_1+\omega)^2}\right]\right\}, \qquad \ldots\ldots(A\,8)$$

in which $\sigma_{\parallel}$ and $\sigma_{\perp}$ are the real parts of the conductivities parallel and perpendicular to the earth's magnetic field H and $\omega=He/m$ (E.M.U.).

The work done by the components of E_1 parallel and perpendicular to H is $W_{\parallel}=(\frac{1}{2}\sigma_{\parallel}E_1^2\cos^2\theta)/N$ and $W_{\perp}=(\frac{1}{2}\sigma_{\perp}E_1^2\sin^2\theta)/N$, where θ is the angle between E_1 and H. When $p_1^2\ll\nu^2\ll\omega^2$ as is the case in these tests, then

$$\sigma_{\perp}\to 0; \qquad \sigma_{\parallel}\to\frac{2}{3}\frac{Ne^2}{m}\left(\frac{\nu}{p_1^2+\nu^2}\right)\left(1+\frac{p_1^2}{p_1^2+\nu^2}\right), \qquad \ldots\ldots(A\,9)$$

and

$$W=AE_1^2, \qquad \ldots\ldots(A\,10)$$

with

$$A=\frac{\cos^2\theta}{3}\cdot\frac{e^2}{m}\left(\frac{\nu}{p_1^2+\nu^2}\right)\left(1+\frac{p_1^2}{p_1^2+\nu^2}\right). \qquad \ldots\ldots(A\,11)$$

The following simplifications of (A 11) are applicable with the restrictions specified:

(*a*) $p_1^2\gg\nu^2$: $A\simeq 2e^2\nu\cos^2\theta/3mp_1^2$;
(*b*) $p_1^2\sim\nu^2$: $A\simeq e^2\nu\cos^2\theta/2m(\nu^2+p_1^2)$;
(*c*) $p_1^2\ll\nu^2$: $A\simeq e^2\cos^2\theta/3m\nu$.

In these special cases W is easily calculated from (A 10); the appropriate value is then used in (22).

REFERENCES

Bailey, V. A., 1937, *Phil. Mag.*, **23**, 744; 1938, *Ibid.*, **26**, 425.

Bailey, V. A., and Martyn, D. F., 1934, *Nature, Lond.*, **133**, 218; *Phil. Mag.*, **18**, 369.

Grosskopf, J., 1938, *Hochfrequenztechn. u. Elektroakust.*, **51**, 18.

Healey, R. H., and Reed, J. W., 1941, *Behaviour of Slow Electrons in Gases* (Amalgamated Wireless—Australasia), § 6.5.

Huxley, L. G. H., 1937, *Phil. Mag.*, **23**, 210, 442.

Huxley, L. G. H., Foster, H. G., and Newton, C. C., 1947, *Nature, Lond.*, **159**, 300.

Tellegen, B. D. H., 1933, *Nature, Lond.*, **131**, 840.

The Diffraction Structure of the Elementary Coma Image

By R. KINGSLAKE

Eastman Kodak Company, Rochester, N.Y. *

MS. received 25 *October* 1947

ABSTRACT. The integrations reported in this paper demonstrate conclusively that the elementary coma image is of the size and shape determined by ordinary geometrical optics, and that the only effect of diffraction is to break up the image into an elaborate fine structure of dots and lines of light. Practically all the light in the image is confined within the triangular space between the principal ray at the tip of the figure and the sagittal focus. Between the sagittal and tangential foci there is a small amount of light broken up into a series of approximately concentric curved bands of darkness and light, centred about the brightest part of the image. Even when the amount of coma is very small, comparable to the Rayleigh Limit, the characteristic shape of a typical coma-image is already making its appearance. The accuracy of the theoretical predictions is fully and completely borne out by the actual photographs of a comatic image.

§ 1. INTRODUCTION

By the aberration "coma" is meant that third-order or Seidel aberration which leads to a wavefront differing from the ideal sphere by an amount P, given by

$$P = -ay(y^2+z^2) \qquad \ldots\ldots(1)$$

where y, z represent the coordinates of a point in the exit pupil of the lens, and a the amount of coma present; a varies linearly with the image height above the lens axis. The centre of curvature of the ideal or reference sphere is assumed to be the ordinary Gaussian extra-axial image point. In the absence of distortion, this is the point where the principal ray of the oblique pencil pierces the paraxial focal plane. The minus sign in (1) indicates that with positive coma, the wave is lagging for points above the middle of the lens.

Since we are discussing one of the Seidel aberrations, we must assume that both the aperture and the obliquity are very small. The longitudinal departure, P, of the wavefront in the exit pupil, relative to the reference sphere, amounts at most to a few wavelengths and is thus infinitesimal compared with y and z.

§ 2. GEOMETRICAL IMAGE

The y and z components of the point where a given ray pierces the paraxial focal plane, relative to the Gaussian image-point, are given by :

$$y' = -l\frac{\partial P}{\partial y}; \qquad z' = -l\frac{\partial P}{\partial z} \qquad \ldots\ldots(2)$$

where l is the longitudinal distance from the exit pupil to the image plane.

Thus

$$y' = al(3y^2+z^2); \qquad z' = 2alyz. \qquad \ldots\ldots(3)$$

* The major part of the computations involved in this paper were made by the author in 1936–37 when he was on the faculty of the Institute of Applied Optics, University of Rochester.

These quantities are best interpreted by expressing the aperture-point position in polar coordinates. Writing

$$y = r\cos\theta; \qquad z = r\sin\theta \qquad \ldots\ldots(4)$$

the components of the ray-point displacement become

$$y' = alr^2(2+\cos 2\theta); \qquad z' = alr^2(\sin 2\theta). \qquad \ldots\ldots(5)$$

The values of y' and z', for a single zone of the aperture, give a double circle of radius alr^2 in the image plane as shown in figure 1, the centre of the circle

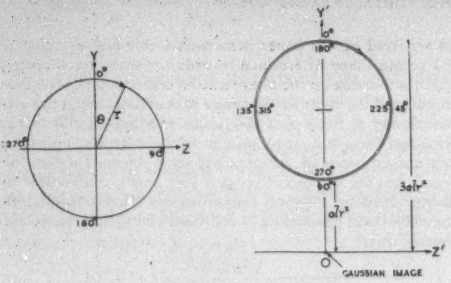


Figure 1. Ray-point in exit pupil and in image plane.

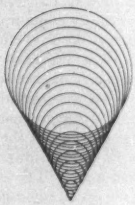


Figure 2. The geometrical coma image.

being at a height $2alr^2$ above the Gaussian image-point. Each elementary zone of the lens forms such a double circle, and the complete family of double circles for the circular lens aperture fit between two common envelope lines, at 60° to each other, which intersect at the Gaussian image or principal-ray point (figure 2).

The top and bottom ("tangential") rays of any zone intersect at the top-most point of the double circle in figure 1, and the front and rear ("sagittal") rays of a zone meet at the lowest point of the double circle. If the height of the tangential image above the principal ray is coma_T, and the height of the sagittal image coma_S, then

$$\text{coma}_T = 3\ \text{coma}_S. \qquad \ldots\ldots(6)$$

§ 3. THE PHASE RELATIONSHIPS OF COMA

The phase difference between light arriving at the Gaussian image-point from the point (y, z) and from the centre of the exit pupil is given by equation (1), since phase $= 2\pi P/\lambda$. If, however, we shift our reference-point to some point vertically above the Gaussian image, we add a further phase difference which can be used to simplify computation of the light distribution. From figure 3 it is seen that the phase difference at D between light travelling along AD and along PD is given by :

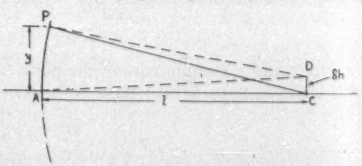


Figure 3. The path difference introduced by a small vertical shift of the reference point.

$$\frac{2\pi}{\lambda}\Big[\text{PD} - \text{AD}\Big] \simeq \frac{2\pi}{\lambda}\left[\frac{y}{l}\cdot \delta h\right] \text{radians.} \qquad \ldots\ldots(7)$$

Hence from (1) and (7), the phase-difference between light arriving at *any* reference-point on the centre-line of the image from (y, z) and from the middle of the aperture is $(2\pi/\lambda)[-ay(y^2+z^2)+(y/l)\delta h]$.

Since for any single zone in the exit-pupil aperture x is negligible compared with y or z, this may be written

$$(2\pi/\lambda)(y/l)[-alr^2+\delta h]. \qquad \ldots\ldots(8)$$

The phase-difference will be zero for *one entire zone* if

$$\delta h = alr^2 \qquad \ldots\ldots(9)$$

and, since alr^2 is equal to the sagittal coma of the zone, we reach the very important conclusion that *each separate zone of the lens forms a perfect equiphase focus at the sagittal image of the zone.* Hence to determine the light distribution in the focal plane, for a complete lens, we divide the lens aperture into a number of circular zones of equal area, and integrate the amplitude contributions of all the zones at each of a number of discrete points in the focal plane. Contours of light intensity can then be drawn through the various points so computed.

§4. THE AMPLITUDE DISTRIBUTION IN THE IMAGE OF A POINT FORMED BY A SINGLE NARROW CIRCULAR ZONE OF THE LENS

This problem is classical, and a good treatment has been given by Martin (1930), who shows that the amplitude at a radial distance t from the centre of the image of a single zone of radius r is given by

$$J_0\left(\frac{2\pi}{\lambda}\cdot\frac{tr}{l}\right). \qquad \ldots\ldots(10)$$

For convenience in computation and in plotting the results, we shall assume that the radius of the lens aperture is unity. We can then introduce a symbol $W=2\pi t/\lambda l$ to represent the radial distance of our integration point R from the centre of a zonal image. By comparing (10) with (7) it may be seen that the physical meaning of W is the phase difference between light reaching the point R from the edge and centre of the (unit) marginal zone of the lens aperture. In the case of coma, there is no path difference for any zone at its sagittal focus, and here W is zero. Thus the amplitude contribution of a zone of radius r, at a radial distance W (radians) from the centre of the zonal image, is given by

$$J_0(Wr). \qquad \ldots\ldots(11)$$

§5. RADIAN MEASURE OF THE AMOUNT OF COMA IN THE LENS

The amount of coma present in the lens may be expressed by a type of radian measure similar to that used in the W notation introduced in §4 by writing

$$B=(2\pi/\lambda l)\,(al)=2\pi a/\lambda. \qquad \ldots\ldots(12)$$

By comparing (1) with (12), it is clear that B represents the phase difference, in radians, between light reaching the Gaussian image-point from the top and centre of the lens aperture.

These two quantities, B and W, may therefore be used to replace coma_S and t respectively, with complete assurance that the units will be properly compatible.

§ 6. THE RAYLEIGH LIMIT FOR COMA

Conrady (1929) has stated that the Rayleigh Limit for Seidel coma is reached when

$$\text{coma}_S = 0{\cdot}5\,\lambda/\sin U'_M \qquad \ldots\ldots(13)$$

which becomes $\lambda l/2$ for a lens of unit radius. Converting this into the B notation, we find that :

$$B\text{-tolerance} = (2\pi/\lambda l).(\lambda l/2) = \pi \text{ radians.}$$

Now Rayleigh based his limit on the fundamental idea that the aberration in a lens will be acceptable if it is just sufficient to cause a diminution of the peak intensity in the image to 80% of its value for a perfect lens. It is therefore of interest to plot the variation of intensity along the centre-line of the image when $B = \pi$ to see how closely Conrady's statement is borne out.

The amplitude at a point on the centre-line of the image, distant Q (radian measure) from the Gaussian image-point, can be found by integrating the zonal contributions given by equation (11). The integration must be taken over a set of infinitesimal zones of equal area, i.e. of equal increment in r^2. In figure 4,

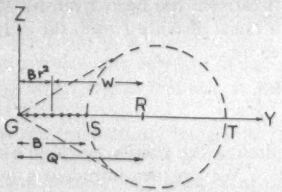


Figure 4. Procedure for integrating the light distribution along the centre-line of the coma figure.

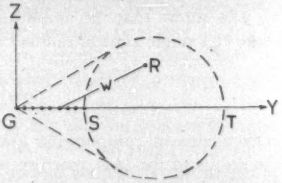


Figure 7. The distance W from a zonal focus to the point of integration, R.

the heavy dots represent the sagittal foci of such a series of zones, equally spaced along the mid-line since the r^2 of the zones are equally spaced; G is the Gaussian image-point, S the sagittal focus of the marginal zone for which $r = 1$, and R the point at which the net amplitude is to be determined by integration. Since the distance GS represents B, and GR represents Q, the distance of R from the equiphase focus of any zone r is $W = (Q - Br^2)$ radians. The total amplitude is thus, by equation (11),

$$\int_0^1 J_0(Wr)d(r^2) = \int_0^1 J_0[r(Q - Br^2)]d(r^2). \qquad \ldots\ldots(14)$$

This equation has been given by Steward (1926 a, b) and Buxton (1926).

A series of values of (14) were calculated for a succession of Q values, for $B = \pi$ and by plotting graphs of the J_0 values against r^2 and integrating with a planimeter, the central section of the coma image could be plotted (figure 5). It is clear that unit intensity can be obtained only when $B = 0$ and $Q = 0$, that

is, at the centre of a perfect Airy disc. It will be seen from figure 5 that Conrady's tolerance is not in accordance with Rayleigh's rule, for the maximum intensity is about 89% of the central Airy disc intensity when $B=\pi$.

Steward's (1926a) data for a central section of intensity of the coma image for the case $B=4$ are shown in figure 6. This case does actually represent the Rayleigh limit, for the peak intensity is now 80%. Conrady's formula (13) above should therefore be emended to read

$$\text{Rayleigh limit for coma}_S = 0{\cdot}64\lambda/\sin U'_M.$$

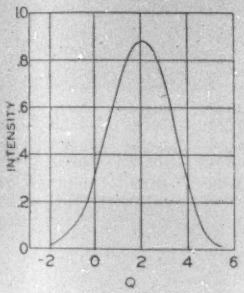


Figure 5. The centre-line section of the coma image, for $B=\pi$.

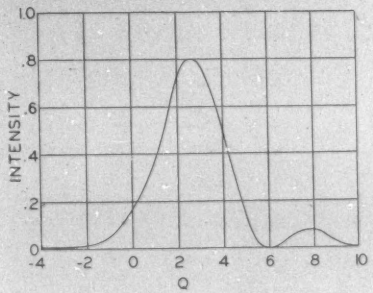


Figure 6. The centre-line section of the coma image, for $B=4$ (*from* Steward).

§7. LIGHT CONTOURS

Light contours for the case $B=4$

To determine the light distribution over the whole image, we assume that the lens aperture has unit radius, and divide the aperture into a number of zones equispaced in r^2. The procedure follows that suggested by Conrady (1919), except that he used Simpson's rule for the mechanical integration instead of a planimeter. A point R is taken in the focal plane (figure 7), and the oblique distances W from R to each of the equiphase foci of the various zones are tabulated. $J_0(Wr)$ is then plotted against r^2 and the area of the graph determined in the usual way with a planimeter. This area is the net amplitude at the point R due to the superposition of the contributions of all the zones in the entire lens. By taking a succession of points R in the image plane, a number of longitudinal and cross sections of the light distribution (in amplitude) can be plotted. The points where these sections cross the desired contour lines of intensity can then be joined with smooth curves.

For the case of $B=4$, the lens was divided into eight zones having radii r equal to $\sqrt{(1/8)}$, $\sqrt{(2/8)}$, $\sqrt{(3/8)}$, . . . 1·0 respectively. The amplitude-contribution graphs were plotted for each of about 150 points in the image plane, having Y values lying between -10 and $+14$ (radian measure) and Z values lying between 0 and 8. As the pattern is symmetrical about the Y axis, only one side need be computed. Each curve was then integrated with a planimeter giving directly the net amplitude at that particular point. The results of the integrations are summarized in table 1.

Table 1. Computed amplitudes for $B=4$

Y \ Z	0	±1	±2	±3	±4	±5	±6	±8
−10	0							
− 9	−0·008							
− 8	−0·002		−0·006					
− 7	0·002							
− 6	0·020		0·024		−0·004	−0·002	−0·012	
− 5	0·044			0·038	0·028	0·014	0·0	
− 4	0·064		0·060	0·070	0·048	0·040	0·016	
− 3	0·078	0·078	0·070	0·052	0·048	0·052	0·036	
− 2	0·104	0·088	0·066	0·038	0·036	0·040	0·052	
− 1	0·218	0·186	0·124	0·052	0·004	0·008	0·034	
0	0·426	0·368	0·218	0·078	−0·016	−0·046	0·014	
1	0·648	0·580	0·376	0·134	−0·028	−0·086	−0·032	
2	0·848	0·734	0·474	0·168	−0·048	−0·118	−0·070	0·060
3	0·888	0·772	0·472	0·158	−0·078	−0·156	−0·110	0·064
4	0·710	0·584	0·342	0·052	−0·142	−0·186	−0·110	0·066
5	0·366	0·294	0·106	−0·106	−0·212	−0·190	−0·090	0·090
6	0·040	−0·018	−0·128	−0·218	−0·236	−0·158	−0·028	0·100
7	−0·208	−0·230	−0·246	−0·246	−0·194	−0·078	0·032	0·110
8	−0·264	−0·258	−0·240	−0·176	−0·088	0·024	0·102	0·088
9	−0·152	−0·148	−0·096	−0·026	0·046	0·106	0·136	0·046
10	0·020	0·012	0·054	0·106	0·130	0·144	0·130	−0·008
11	0·134	0·118	0·152	0·160	0·136	0·100	0·058	−0·052
12	0·130	0·128	0·128	0·116	0·070	0·038	−0·018	
13	0·076	0·050	0·052	0·020	−0·020			
14	−0·034	−0·044		−0·052				

The centre-line amplitudes are a recomputation of those given by Steward and shown in figure 6.

A series of longitudinal and transverse sections were then plotted from the above figures, and contour points were taken from the section graphs at amplitude values $\sqrt{I}$ corresponding to intensities $I=0$, 2, 4, 6, 8, 10, 20, 30, 40, 50, 60, 70, 80%.

The resulting contour map of the light distribution is seen in figure 8.* In this figure, the zero-intensity loci are shown dotted, and the outline of the geometrical image is included.

Light contours for the case $B=3$

It has recently been brought to my attention that the intensity contours for the case of $B=3$ have been very fully worked out by Nijboer (1946). This plot is reproduced in figure 9 as it may not be readily available.

Light contours for the case $B=6{\cdot}14$ *radians* (352°)

This case has been integrated by Martin (1922 a), and the central part of the pattern plotted. It was later verified photographically by the use of a specially designed microscope objective having the desired amount of coma (1922 b).

Light contours for the case $B=20$

For this case, 16 zones were taken of radii $\sqrt{(1/16)}$, $\sqrt{(2/16)}$, . . . 1, respectively, the equiphase-points of the successive zones being, therefore, 1·25 (radian) units apart. The amplitude-contribution graphs were plotted for some 190 points in the focal plane, and each graph was integrated as before. The resulting net amplitudes are tabulated in table 2. The contour pattern (figure 10) was plotted as before by drawing a series of longitudinal and transverse sections and marking the points for the amplitudes $\sqrt{I}$ corresponding to $I=0{\cdot}5$, 1, 2, 4, 6, 8, 10%.

* This pattern differs somewhat from that given by Steward (1926 a, p. 156); there are also small differences between the amplitude distribution along the centre-line of the pattern as given in table 1 and as given by Steward.

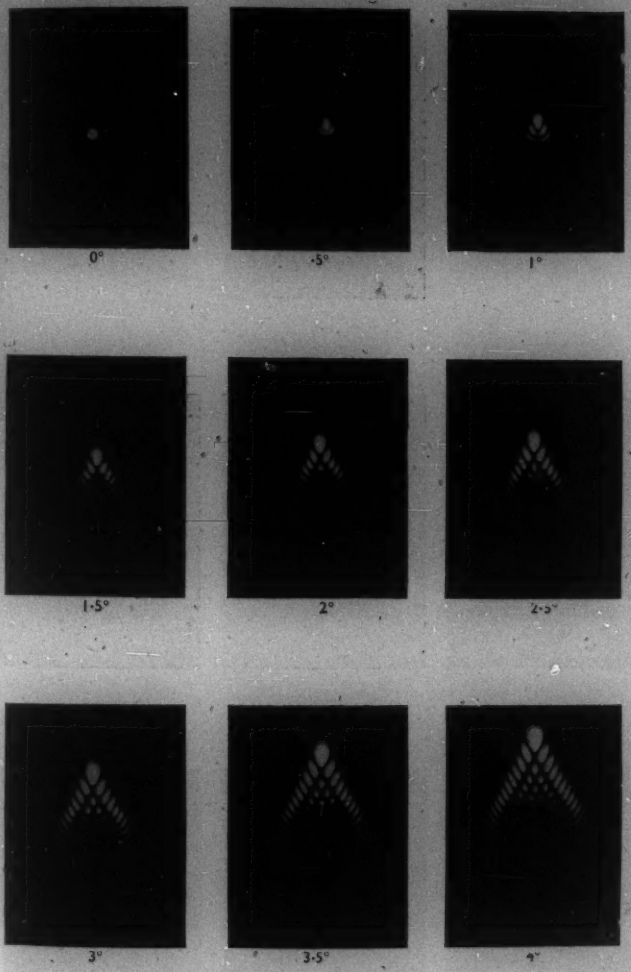


Figure 13. Photographs of a comatic star image showing progressively increasing amounts of coma.

To face page 152

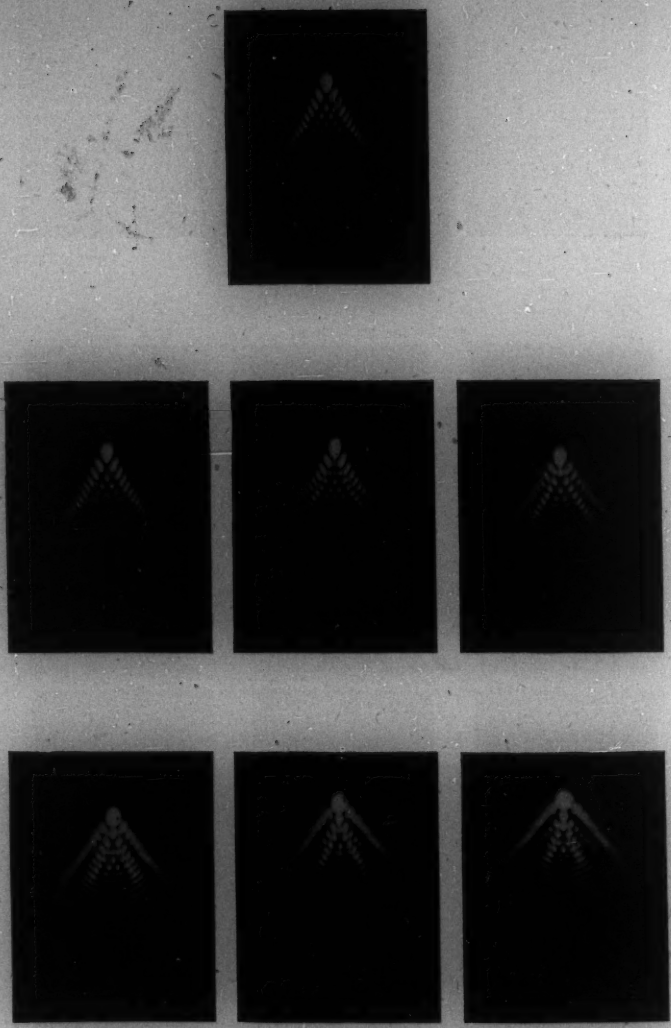


Figure 14. Photographs of a comatic star image showing the effects of progressively increasing departures from the focal plane.

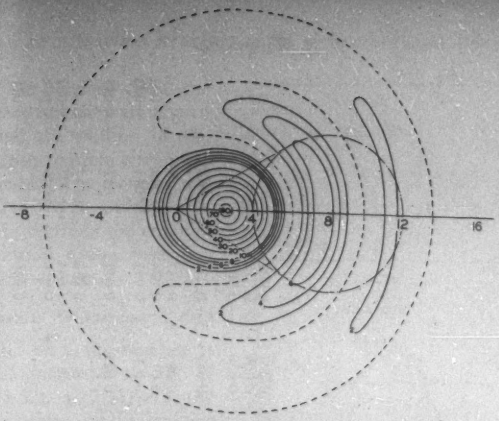


Figure 8. Contours of intensity for $B=4$

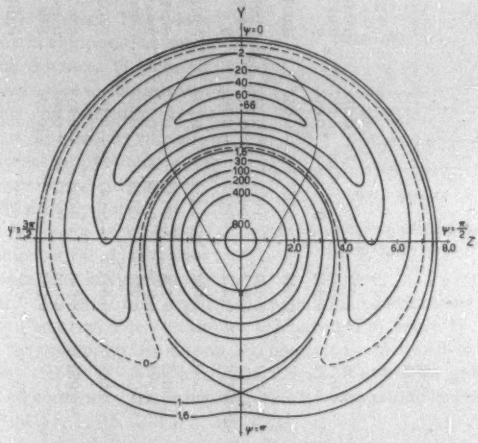


Figure 9. Contours of intensity for $B=3$ (*from* Nijboer).
Reproduced by kind permission of the Elsevier Publishing Co., Inc.

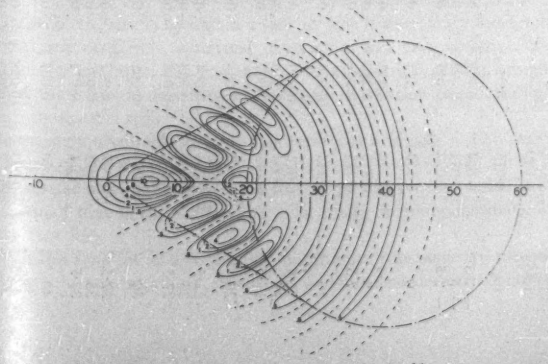


Figure 10. Contours of intensity for $B=20$.

Table 2. Computed amplitudes for $B=20$

Y \ Z	0	±2	±4	±6	±8	±10	±12	±14	±16	±18	±20	±22
−15	−0·002											
−12·5	0·010											
−10	0·016	0·011										
− 7·5	0·024	0·012	0·014									
− 5	0·041	0·021	0·016	0·010								
− 2·5	0·072	0·067	0·044	0·014								
0	0·133	0·115	0·069	0·027	0·012							
2·5	0·226	0·185	0·082	0·018	−0·021		0·013	0·008				
5	0·318	0·240	0·093	−0·014	−0·032		0·017	0·011				
7·5	0·315	0·204	0·016	−0·074	−0·042	0	0·016	0				
10	0·245	0·074	−0·130	−0·160	−0·046	0·051	0·053	−0·004				
12·5	0·073	−0·079	−0·220	−0·119	0·044	0·079	0·021	−0·007	−0·007	−0·002		
15	0·008	−0·119	−0·213	0·014	0·177	0·068	−0·037	−0·009	0·004	0		
17·5	0·145	0·007	−0·056	0·158	0·211	−0·025	−0·109	0	0·030	−0·008		
20	0·142	0·086	0·067	0·160	0·058	−0·153	−0·093	0·071	0·039	−0·032	0·008	
22·5	−0·004	−0·004	0·016	−0·012	−0·128	−0·142	0·032	0·106	−0·003	−0·057	0·018	
25	−0·064	−0·063	−0·065	−0·093	−0·076	0·039	0·128	0·042	−0·073	−0·034	0·046	−0·002
27·5	−0·015	−0·016	−0·009	0·029	0·095	0·128	0·033	−0·096	−0·070	0·041	0·044	−0·023
30	0·089	0·091	0·090	0·088	0·042	−0·042	−0·105	−0·072	0·035	0·071	−0·012	−0·041
32·5	−0·007	−0·007	−0·033	−0·076	−0·098	−0·088	−0·019	0·067		−0·016	−0·057	0
35	−0·072	−0·070	−0·069		0·014	0·072	0·084	0·044		−0·064	−0·007	0·037
37·5	0·071	0·074	0·086		0·065	0·032	−0·026	−0·073		0·016	0·046	0·004
40	0·010	0·007	−0·014			−0·077	−0·054	0·002		0·046		−0·041
42·5	−0 073	−0·058	−0·057			0·028	0·054					−0·004
45		0·069	0·060			0·033	−0·009					
47·5		−0·016	−0·018				−0·040					
50			−0·021									

As in the previous cases, the loci of zero intensity are shown dotted, and the outline of the geometric coma pattern is included for reference. The pattern was computed to $Y=50$, which is just beyond the centre of the circle in the geometrical image belonging to the marginal zone.

Light contours for the case $B=40$

For this case, the number of zones was increased to 32, of radii $\sqrt{(1/32)}$, $\sqrt{(2/32)}$ etc. the equiphase-points of successive zones being 1·25 (radians) apart as for the $B=20$ case. This was a convenience in computation since the oblique distances Q from all the equiphase foci to the integration points had already been evaluated. The labour of computing the net amplitude at each integration point was considerable, as it involved computation of 32 four-digit products of r and Q, determination of the J_0 curve for each product and integration by planimeter. To show how rapidly the amplitude contributions alternate from one zone to another, a few examples are included in figure 11.

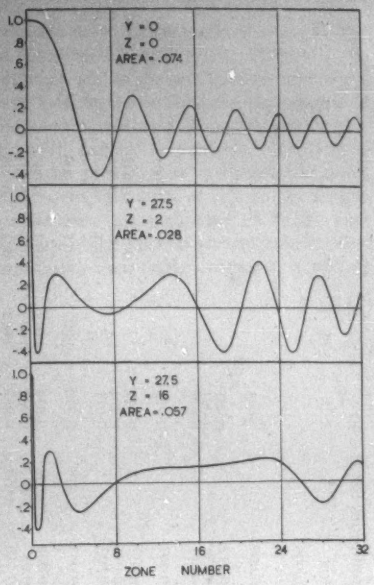


Figure 11. Some typical amplitude-contribution curves for the case $B=40$.

In all, the net amplitudes at some 256 points in the image plane were computed in this way (table 3). The longitudinal and cross section graphs were then plotted, and contours were drawn (figure 12) at amplitude levels corresponding to intensities 0·5, 1, 1·5, 2, 2·5, 3, 3·5, 4%.

§8. PHOTOGRAPHIC CONFIRMATION OF THE COMPUTED PATTERNS

The problem here is to obtain a lens in which there is a considerable amount of coma and negligible spherical aberration and astigmatism. Such a lens might be, for example, the rear half of a symmetrical system designed to give excellent anastigmatic definition at unit magnification, with parallel light entering through the central stop.

Fortunately a suitable lens of this type was available. The rear half had a focal length of about 8 inches and an aperture of f/5·6, the field out to 5° from the axis being flat and highly corrected for astigmatism. The spherical aberration was so small that even at f/5·6 the axial image in monochromatic light was a perfect Airy disc.

With this lens set at 4° obliquity, an attempt was made by adjusting the iris diaphragm to reproduce as closely as possible the computed pattern for the

Table 3. Computed amplitudes for $B=40$

Y \ Z	0	±2	±4	±6	±8	±10	±12	±14	±16	±18	±20	±22	±24	±26
−5	0·028						−0·006							
−2·5	0·051	0·030												
0	0·074	0·079	0·043	0·021	0·009	0	−0·008		0·003					
2·5	0·129	0·115	0·073	0·036	0·009	−0·004	−0·012		0·005					
5	0·178	0·150	0·093	0·030	−0·008	−0·019	−0·011		0·001					
7·5	0·200	0·163	0·090	0·006	−0·027	−0·026	−0·011		0·011					
10	0·181	0·137	0·038	−0·049	−0·072	−0·041	−0·007		0·021		0	−0·001		
12·5	0·135	0·075	−0·030	−0·101	−0·088	−0·032	0·014		0·012		0·001	−0·002		
15	0·054	−0·005	−0·109	−0·139	−0·072	0·010	0·043		−0·001		−0·006	−0·005		
17·5	0·015	−0·053	−0·139	−0·136	−0·016	0·063	0·067		−0·011		−0·001	0		
20	0·020	−0·037	−0·106	−0·049	0·074	0·132	0·063	−0·010	−0·030	−0·007	0·011	0·008		
22·5	0·082	0·021	−0·041	0·026	0·139	0·116	0·008	−0·063	−0·039	0·014	0·023	0·006		
25	0·142	0·068	0·002	0·077	0·131	0·050	−0·087	−0·087	0·002	0·041	0·013	−0·003		
27·5	0·100	0·028	−0·027	0·033	0·052	−0·054	−0·130	−0·052	0·057	0·062	−0·002	−0·026		
30	0·017	−0·041	−0·089	−0·024	−0·009	−0·096	−0·088	0·044	0·101	0·032	−0·036	−0·026		
32·5	0·004	−0·056	−0·091	−0·005	0·009	−0·068	0·014	0·136	0·079	−0·043	−0·058	−0·002	0·025	
35	0·054	−0·016	−0·033	0·073	0·073	−0·009	0·061	0·118	−0·018	−0·115	−0·021	0·051	0·022	
37·5	0·119	0·044	0·014	0·093	0·070	−0·017	0·029	0·025	−0·095	−0·089	0·064	0·078	−0·022	
40	0·042	0	−0·002	0·035	−0·005	−0·063	−0·025	−0·025	−0·076	0·019	0·118	0·031	−0·071	−0·017
42·5	−0·059	−0·062	−0·042	−0·036	−0·052	−0·049	−0·002	0·001	0·013	0·087	0·070	−0·063	−0·074	0·033
45	−0·012	−0·016	−0·014	−0·010	0·004	0·036	0·056	0·038	0·042	0·040	−0·041	−0·091	−0·004	0·080
47·5	0·035	0·039	0·041			0·054	0·023	−0·008	−0·013	−0·042	−0·077	−0·011	0·079	0·046
50	0·021	0·015	0·018			−0·035	−0·047			−0·033	0·003	0·069	0·053	−0·042
52·5	−0·034	−0·035	−0·037			−0·041	−0·022			0·033	0·056	0·027		−0·017
55	−0·018	−0·027	−0·006				0·042				−0·004			0·013
57·5	0·048													0·060
60	0·004													

case $B=40$. Then, since coma is proportional to image height, it was expected that each degree of obliquity would represent 10 units of B. However, this operation proved to be surprisingly difficult, and in the actual series of enlarged photographs of the star image at every half degree of obliquity, reproduced in figure 13, it will be seen that the computed patterns for $B=4$, 20, and 40 very closely resemble the photographs at $\frac{1}{2}°$, 2°, and $3\frac{1}{2}°$ respectively. The resemblance between the computations and the photographs is remarkably close, even to the shape of the individual little spots and bands of light in the images.

The original star-image photographs were made through an 8 mm. apochromatic microscope objective on ordinary 16 mm. motion-picture panchromatic film, in sodium D light, the negatives being then further enlarged in making the prints.

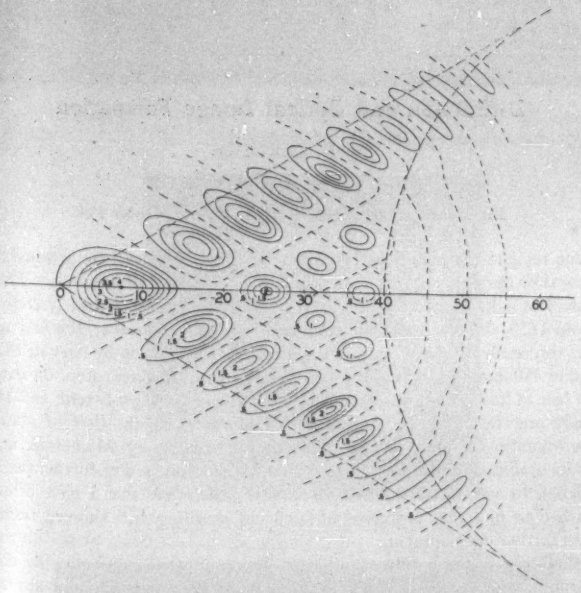


Figure 12. Contours of intensity for $B=40$.

§ 9. OUT-OF-FOCUS COMA IMAGES

The problem of computing the light distribution in an out-of-focus coma image is very difficult, as the convenient fact that each zone forms a sharp equiphase focus at its sagittal image can no longer be utilized. Consequently, use was made of the lens system mentioned in the last paragraph to provide this information. The lens was tilted to $4\frac{1}{2}°$ obliquity (corresponding approximately to $B=50$), and a succession of star-image photographs were made for a series of focal settings at successive intervals of 0·1 mm. out of focus. It was found to be immaterial whether the image plane was chosen within or beyond the best

focus plane, thus again confirming the very high degree of correction of all the aberrations other than coma. These out-of-focus star images are shown in figure 14.

REFERENCES

BUXTON, A., 1926, *Proc. Opt. Conv.*, II, 771.
CONRADY, A. E., 1919, *Mon. Not. R. Astr. Soc.*, **79**, 577 ; 1929, *Applied Optics and Optical Design*, I, 395.
MARTIN, L. C., 1922 a, *Mon. Not. R. Astr. Soc.*, **82**, 310 ; 1922 b, *Trans. Opt. Soc.*, **24**, 1 ; 1930, *Applied Optics*, I, 95.
NIJBOER, B. R. A., 1946, Part of thesis embodying the coma analysis—see Bouwers, A., *Achievements in Optics*, Monographs on the Progress of Research in Holland (New York and Amsterdam : Elsevier Publ. Co., Inc.).
STEWARD, G. C., 1926 a, *Phil. Trans. Roy Soc.* A, **225**, 153 ; 1926 b, *Proc. Opt. Conv.*, II, 791.

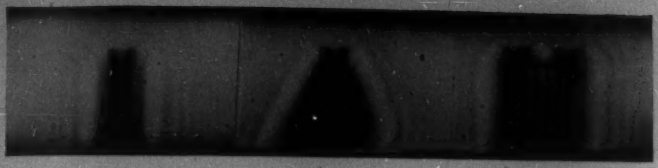
Diffraction and Optical Image Formation

By F. ZERNIKE
University of Groningen

The Thomas Young Oration, delivered 24 September 1947

IF one reviews the historical development of this subject, one is struck by the very slow development of the fundamental concepts and methods. The wave theory of light seemed not only well established, but even nearly completed, by about 1820, after the work of Young and Fresnel. However, the first application to the resolving power of optical instruments was made by Airy in 1835, the second by Helmholtz and by Abbe forty years later. The next step, the extension to the case of lens errors, was not made until about 1900 by Strehl, and 1920 by Conrady and by Richter. A somewhat different branch, that of coherence, begun hesitatingly by Verdet in 1860 and developed by Michelson in 1890, found its practical application about 1920 at Mt. Wilson, and its further theoretical foundation by van Cittert (1934). You will understand that I have often been astonished to find that problems of such old standing still showed themselves open to further development.

I shall treat of three different, though interconnected, subjects: the coherent background, the degree of coherence and the diffraction theory of aberrations. As an introduction, I would demonstrate a few diffraction experiments. By the aid of a small arc lamp, condenser and vertical slit I throw the shadow of a thin vertical needle on the screen. In order to make the details visible through the whole room, I enlarge the shadow, in the horizontal direction only, by a short focus *cylindrical* lens. The external fringes are clearly seen, but also the internal fringes, which Thomas Young explained by interference between the two beams diffracted at the edges (figure 1 (*a*)). With a wedge-shaped needle (figure 1 (*b*)) the different behaviour of the two kinds with increasing thickness of the needle is seen at a glance. The internal fringes soon become too dark to be seen. Yet with another thicker needle they reappear very clearly, a dark fringe at the centre (figure 1 (*c*)). I have used a simple trick here: the needle is double, through



(a) (b) (c)

Figure 1.

Figure 2.

Figure 3. Diffraction figure of a slit, *above* without background, *middle* with coherent background, *below* background alone.

Figure 4. F. G. Pease at the eyepiece of his 15-metre stellar interferometer

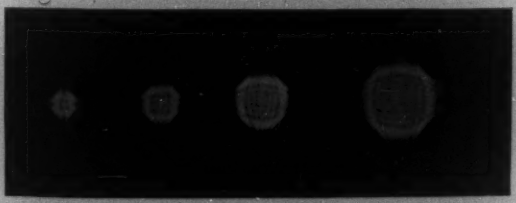


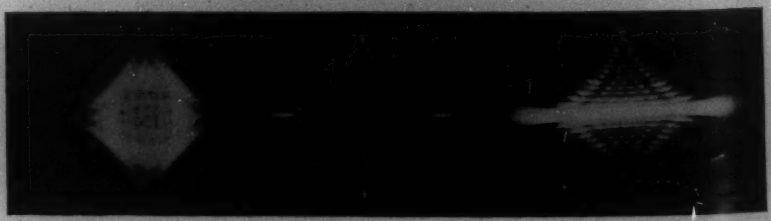
Figure 5. Astigmatism patterns for $\beta=8$, 15, 26 and 45 respectively.

To face page 158

Figure 6. Astigmatism for $\beta=1$, *left* without background, *middle* same with background, *right* background alone.

Figure 7. Coma patterns with $\beta=0{\cdot}7$, $2{\cdot}5$, 10, 20 and 50 respectively.

Figure 8. Astigmatism with $\beta=12$, without background, with background in phase and with background one quarter behind.



(*a*) (*b*) (*c*)

Figure 9. Astigmatism, $\beta=17$, (*a*) complete pattern, (*b*) edge effect alone, (*c*) pattern at focal line.

the narrow slit between some light passes and by diffraction is spread over the dark shadow. Evidently, the background thus created is *coherent* with the faint light of the fringes. Therefore the latter are effective here with their *amplitude*. A simple calculation shows that the fringes are much more easily seen in this way.

In order to obtain a similar effect with the light bent into the shadow at a single edge, I use a different artifice: the shadow-throwing screen is made slightly transparent. A number of fringes is now seen in the shadow (figure 2). In general it is found advantageous to combine both methods. Thus the Fraunhofer diffraction image of a slit was thrown on a coherent background in the following way: The slit is covered by a plane-parallel glass covered with a transparent layer of a strongly absorbing metal (silver or aluminium). In this layer a narrow scratch is made exactly in the centre of the slit. The extra light through this scratch gives rise to a much broader, but not too much weaker, diffraction image, the central fringe of which may cover the whole diffraction image of the slit (figure 3).

In these cases one can in different ways also change the phase of the background at will. The result is therefore that the amplitudes and phases at various points of a diffraction image can be observed experimentally by aid of the coherent background.

In this method there is further a clear parallelism with theory. In order to define clearly the meaning of the phase difference between points of a diffraction image, the theorist must be sure to introduce a *surface of reference*, which may be plane, cylindrical, spherical, etc. in a more or less arbitrary way. Indeed, controversies have sometimes arisen through neglect of this. In the same way the experimenter must introduce an *auxiliary coherent wave*, which may radiate from a more or less arbitrary point or line.

The question of *partial coherence* started with Verdet, who asked at what distance apart two points on a screen illuminated by the sun would still have coherent vibrations. He found that this was determined by the apparent diameter of the sun, the actual distance being less than 1/20 millimetre. Let us imagine an experimenter who wants to verify this. He will take a piece of tinfoil and prick very small holes in it, in pairs of various distances. Through each hole a cone of light will pass and adjacent cones will overlap and show interference fringes. But if they do not, shall we call the adjacent cones *incoherent*? I agree, but I must warn you that this is quite a daring step, undertaken only recently. Indeed, opticians have been very cautious, calling vibrations incoherent only when they came from different sources, thus making sure that their haphazardly changing phases would be statistically independent. Well, as soon as we take the more daring point of view, a new opportunity presents itself, namely to call vibrations partially coherent when they give fringes of lower visibility and to define their *degree of coherence* γ to be equal to the number between 0 and 1 which expresses their visibility:

$$\text{visibility} = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \text{degree of coherence } \gamma.$$

This new concept of degree of coherence leads to various remarkable results. For our sunlit screen γ vanishes at a distance of 0·07 mm., but at larger distances it rises again and goes up and down many times. More exactly, its course is

equal to that of the amplitude of the diffraction image of a star formed by a telescope objective of angular aperture equal to the sun's diameter (Airy disc with rings). It would carry us too far from our subject to discuss the application of the same theorem (discovered by van Cittert 1934) to the rôle of a microscope condenser or of a condenser in front of a spectroscope slit.

Another general theorem, on the contrary, is directly connected with image formation. It is the theorem about the propagation of γ, which states that a knowledge of the distribution of intensity, of degree of coherence and of relative phases in any surface intersecting a beam of light enables us to calculate the same quantities at a following or at a preceding surface. For instance, the image that will be formed in a photographic camera—i.e. the distribution of intensity on the sensitive layer—is present in an invisible, mysterious way in the aperture of the lens, where the intensity is equal at all points, namely in the distribution of γ in this aperture. And if you ask for the mathematical connection between the two: one is the Fourier transform of the other. This was probably known, in an incomplete form, to Michelson in 1890, but for lack of the requisite term and even of the requisite concept, he could not adequately express it. In such cases one feels the truth of E. Mach's statement that science serves to economize thinking. Indeed a single term may stand for a whole theory, may convey its ideas, theorems and concepts.

Michelson (1890) had suggested that it is possible to obtain the apparent diameters of stars from the visibilities of interference fringes, or, in our terminology, to find them from a determination of the degree of coherence as a function of distance apart. In the practical execution of this idea at the Mt. Wilson Observatory, our minute holes in the tinfoil were represented by two eight-inch mirrors under 45° which were movable along a six-metre steel beam, mounted across the opening of the 100-inch telescope. The mirrors reflect the light from a star towards the centre of the beam. Two fixed mirrors mounted there throw the light into the telescope. The observer first sees two star discs which he brings into coincidence by adjusting the mirrors. He then estimates the visibility of the fringes that appear and repeats this for various distances of the first mirrors. F. G. Pease, the designer of the 100-inch telescope, is to be credited for most of this work. It may be said that he devoted much of his time during the last fifteen years of his life to measuring degrees of coherence—without knowing it. All the same he measured various star-diameters and "resolved" Mizar, before only known as a spectroscopic double star (distance 0·011″). Figure 4 shows Dr. Pease at the eyepiece of a specially constructed 15-metre instrument. Perhaps future astronomers will build larger instruments of this kind, say of 50 or 100 metres, which will indirectly show details down to one thousandth of a second of arc.

This indirect method of studying celestial objects may well be compared with the study of crystal structure by x-ray diffraction. There also the synthesis from the Fourier transform to the image cannot be obtained by the direct optical method, that is, the image cannot be seen but must be calculated.

I shall dwell somewhat longer on my last subject, the rôle of diffraction in image formation in the presence of lens errors. The prevailing attitude among opticians was even recently the following. Diffraction causes a certain unavoidable deterioration of the ideal point image, which may be expressed by the radius a of the Airy disc. Any aberration present will also give a certain diffusion,

expressible by a "radius" b of the corresponding geometrical pattern. Both together may then be assumed to give a diffusion $a+b$. There is one thing in favour of this crude estimate: the tolerances for aberrations deduced from it are very much on the safe side.

My own work on the change of the Airy pattern caused by small aberrations was started in 1934 and completed by Nijboer in his thesis of 1942. A short survey of the results must suffice here. To begin with, the aberration will of course be expressed, not in terms of light rays and their intersection with the receiving plane, but as deviations of the wave surface in the exit pupil from the ideal spherical form. Let these deviations be represented by $V(y, z)$ as a function of the plane rectangular coordinates y, z in the circular opening, or as $V(r, \phi)$ in polar coordinates, such that $r=1$ at the edge. Secondly, this characteristic function V is developed in a series of *orthogonal polynomials*,

$$V(r, \phi)=\Sigma\, \beta_{nm} R_n^m(r) \cos m\phi, \qquad \ldots\ldots(1)$$

which were specially constructed for the purpose. In the receiving plane with polar coordinates ρ, ψ the resulting diffraction image has then, to a first approximation, the amplitude

$$A(\rho, \psi)=2\rho^{-1}\{J_1(\rho)-\Sigma\, i^{n+1}\beta_{nm} J_{n+1}(\rho) \cos m\psi\}. \qquad \ldots\ldots(2)$$

Nijboer also gives the general form of the terms up to the fourth powers of the coefficients β.

At the centre of the pattern, the point of maximum amplitude for small errors, the result is

$$A_0=1-\Sigma\, \beta_{nm}^2/4(n+1). \qquad \ldots\ldots(3)$$

The intensity at the centre, A_0^2, is a good measure of the quality of the image (Strehl's "definition-brightness"). A diminution of 10% may well be tolerated, $A_0=0{\cdot}95$. Suppose, for instance that there is only ordinary spherical aberration, β_{40} must then be less than 1, whereas considered geometrically, this value would give a circle of least confusion of radius 6, or 3·1 times the radius of the Airy disc. Another remarkable advance lies in the fact that there are no mixed terms in (3), so that a higher aberration cannot be improved by small amounts of a lower one. In other words, the *balancing* of aberrations has been completely attained by our introduction of orthogonal polynomials.

The method described was thus fully successful for small errors. Dr. Nijboer found it increasingly difficult for larger aberrations. I can show a slide with the pattern for astigmatism with $\beta_{22}=4$. The formula for this case fills a whole page. As it appeared hopeless to get any further theoretically, especially to get an insight into the gradual transition to the geometrical pattern for increasing errors, we turned to experiment. Let me give a few details of the way we obtained pure third order errors.

Astigmatism was obtained with a symmetrical biconvex lens of 1 m. focus and magnification one, thus excluding coma, turned into an oblique position through measurable angles up to 15°. The circular diaphragm remained fixed, perpendicular to the beam. The astigmatism is proportional to the square of the angle.

Coma was obtained from an ordinary achromatic telescope objective by shifting its components laterally in opposite directions. The coma is proportional to the shift and to the third power of the aperture.

Spherical aberration was obtained from a single meniscus lens in a reversed position, the amount being changed by changing the aperture. It is proportional to the fourth power of the aperture.

In all cases green mercury light was used and the diffraction pattern enlarged 5–10 times by an auxiliary lens. The largest amounts of aberration used could always be determined experimentally from geometrical optics (axial distance of focal lines etc.). Let me show a few examples of astigmatism and of coma (figures 5, 6 and 7). Mr. Nienhuis, who made these experiments and whose thesis is to appear in a few months, now proceeded to the experimental investigation of the formation of these patterns.

As an example I take the largest coma pattern. The wave retardation, expressed in radians, is in this case

$$V(r, \phi) = 3\beta r^3 \cos\phi = 3\beta y(y^2 + z^2), \qquad \ldots\ldots(4)$$

there being no need here for the polynomial. Let us start from the geometric optical pattern. As is well known, each zone of the lens aperture gives rise to a circle with displaced centre and which is *described twice*. In fact, the optical path to the point (η, ζ) in the receiving plane becomes

$$V(y, z) - y\eta - z\zeta, \qquad \ldots\ldots(5)$$

and the point of intersection of the ray is found by equating the derivatives of (5) to zero:

$$\eta = \partial V/\partial y, \quad \zeta = \partial V/\partial z, \qquad \ldots\ldots(6)$$

which in our case reduces to $\eta = 3\beta(2r^2 + r^2\cos 2\phi)$, $\zeta = 3\beta r^2 \sin 2\phi$. Therefore two rays from diametrically opposite points of a zone intersect in the same point η, ζ and must show interference. Substituting (6) into (5), the path difference with the principal ray becomes generally $V - y\,\partial V/\partial y - z\,\partial V/\partial z = V - 3V = -2V$, or twice this amount between the two interfering rays. Nienhuis finds that the observed appearance of the interference fringes is quantitatively explained in this way. In the lower part of the coma pattern, near the tip, the lower parts of large circles will overlap with the upper parts of much smaller ones. The two crossed systems of fringes give rise to the diamond pattern observed.

One further detail can better be illustrated on the astigmatism pattern. In this case the geometrical pattern is to be deduced in the same way from $V(y, z) = \beta r^2 \cos 2\phi = \beta(y^2 - z^2)$ giving $\eta = 2\beta y$, $\zeta = -2\beta z$, that is, all zones give concentric and proportional circles which are described in opposite directions. Therefore there is no overlapping here and we should expect an evenly illuminated circle instead of the observed four-pointed star. In the same way as above, we further obtain for the path difference in the receiving plane,

$$V - 2V = -V = -\beta(y^2 - z^2).$$

This means that our pattern, on interfering with an auxiliary spherical wave, should show the same hyperbolic fringes, with reversed sign, that would be found in the lens aperture in a Twyman interferometer. The method of the coherent background realizes this; figure 8 shows the result.

The four-pointed star may be explained by the diffraction at the diaphragm-edge. As is well known, the Fresnel diffraction at an edge of any form may be ascribed to rays emerging from the edge and spreading from the undiffracted ray only in directions perpendicular to the edge. In our case we must therefore

expect diffraction streaks issuing from each point on the circle of least confusion. As this circle is described in the opposite direction, the streaks will turn *against* the radius vector. The geometrical problem is easily solved, the streaks will envelop an asteroid, the equation to which will be $\eta^{\frac{2}{3}}+\zeta^{\frac{2}{3}}=(4\beta)^{\frac{2}{3}}$. This was tested experimentally by inserting a metal disc in the diaphragm opening, leaving free only a narrow annulus. The asteroid pattern then appeared, unobstructed by the intense geometrical pattern (figure 9 (*b*)). The reverse is also possible: by throwing the shadow of a small circular opening on the lens aperture, the latter may be illuminated with an intensity decreasing towards the edge and vanishing at the edge itself. The edge effect was indeed absent in this case, only a simple circular pattern remaining. Such experimental tricks are not even necessary. It is found geometrically that the asteroid edge-pattern must remain unchanged, at least in form, when the focus is changed. It therefore comes out much clearer when the receiving plane is placed at one of the focal lines (figure 9 (*c*)).

After these experiments the question arose: must we rest content with this solution of the problem? Evidently the problem is in essence a mathematical one; we have no reason to doubt the validity of the relatively simple diffraction integral, which was also the starting point for the development in case of small errors. The real difficulty was to find an *asymptotic expansion* for large wave numbers. Only a few months ago my assistant N. G. van Kampen attacked this problem again and found the solution. The "method of stationary phase" used goes back to Stokes and Kelvin. Its mathematical elaboration was well known to us, as it is due to my colleague J. G. van der Corput (1936). However, van Kampen had to extend it to two variables. Let me give a brief summary of the mathematical formulation. The problem is to develop an integral of the general form

$$\int_D kg(y,z)e^{ikf(y,z)}\,dy\,dz$$

into an asymptotic series for large values of the parameter k, the integral being extended over the domain D, the boundary of which consists of a finite number of analytical curves.

It is found that the ever increasing rapidity of fluctuation of the exponential causes the series to depend on the behaviour at a limited number of *decisive points*, which are of three kinds: (*a*) internal points at which the argument of the integrand is stationary, i.e. $\partial f/\partial y=\partial f/\partial z=0$, (*b*) boundary points at which the argument is stationary along the boundary, i.e. $\partial f/\partial s=0$, (*c*) corner points, or boundary points where two analytical curves join.

In the neighbourhood of any internal decisive point (y_0, z_0), the exponential may be partially developed into a power series

$$\exp(ikf_0)\exp\{ik(\alpha_{11}y^2+2\alpha_{12}yz+\alpha_{22}z^2)\}\{1+\ldots ik\alpha_{31}y^3+ik\alpha_{41}y^4\}$$

and this is integrated term by term between limits and $-\infty$ $+\infty$. The resulting integrations are easily performed, they give a series with principal term

$$\pi g_0\, e^{ikf_0}\begin{vmatrix}\alpha_{11} & \alpha_{12}\\ \alpha_{12} & \alpha_{22}\end{vmatrix}^{-\frac{1}{2}} \qquad \ldots\ldots(7)$$

with additional details about signs into which I shall not enter. The following terms are of order k^{-1}, k^{-2} etc. In a similar way, each decisive boundary point gives a series beginning with $k^{-\frac{1}{2}}$, $k^{-\frac{3}{2}}$ etc. and each corner point terms of order

k^{-1}, k^{-2} etc. The contributions of all decisive points must finally be added. Of course it is no easy matter to prove the mathematical validity of the whole procedure, but Professor van der Corput has just now succeeded in it.

Now, applying this to our diffraction integral, we may ordinarily put the amplitude g equal to one and $f = V(y, z) - y\eta - z\zeta$, and consider a circular boundary. The internal decisive points are then to be found from (6) and the principal term (7) is that of geometrical optics, with phases and interference taken into account. Even the intensity agrees exactly, the square of the last factor of (7) corresponding to the concentration of light rays by the curvature of the wave surface.

The second term comes from the boundary points, it corresponds to the beams diffracted by the edge. But here the theory gives more than the experimental treatment, which was not able to predict the intensity. Of course the new development also gives more than these two terms and these further terms could not be found in another way. We are now endeavouring to fill the gap between small and large errors by calculating some intermediate case, say astigmatism with $\beta = 10$, from both sides.

It is especially the general insight, however, which gains very much by the discovery of the asymptotic development. It shows that physical intuition combined with experimental ability may go far towards elucidating the main characteristics of phenomena, but that only an adequate mathematical treatment can give a satisfactory final solution.

REFERENCES

BIJL, J., 1938, *Nieuw Arch. Wisk.* (2), **19**, 63.
VAN CITTERT, P. H., 1934, *Physica*, **1**, 201.
VAN DER CORPUT, J. G., 1936, *Compos. Math.*, **3**, 328.
MICHELSON, A. A., 1890, *Phil. Mag.*, **30**, 1.
NIJBOER, B. R. A., 1942, *Thesis, Groningen.*
PEASE, F. G., 1931, *Ergebn. exakt. Naturw.*, **10**, 84.
ZERNIKE, F., 1938, *Physica*, **5**, 785 ; 1946, *C.R. Réunions Opticiens*, in press.

The Structure and Growth of PbS Deposits on Rocksalt Substrates

BY A. J. ELLEMAN AND H. WILMAN
Applied Physical Chemistry Laboratory, Imperial College, London

MS. received 28 *November* 1947

ABSTRACT. The structure of PbS deposits condensed from the vapour *in vacuo* on to {001}, {110}, {111} and {443} rocksalt faces has been investigated by electron diffraction. The results suggest that the deposit atoms take up positions of least potential energy relative to the substrate, as far as is permitted by the disturbing effects of collisions of incident atoms with the initial deposit crystal nuclei, and by the limited surface mobility of the deposited atoms over the substrate. This view is also supported by the nature of the changes in crystal orientation which occur when initially random deposits are heated *in vacuo*.

§1. INTRODUCTION

LEAD sulphide is a semiconductor which has important applications owing to its photoconductive properties and especially its electron-emission sensitivity to infra-red radiation. Outlines of the properties of sensitive deposits, and of their production by sublimation *in vacuo* or low pressure oxygen have been published by Lee and Parker (1946), Starkiewicz *et al.* (1946), and Sosnowski *et al.* (1947 a, b).

A recent investigation (Wilman 1948) of such deposits by electron diffraction established that lanarkite, $PbO.PbSO_4$, was formed as an oxidation product in the sensitization of the PbS by sublimation and baking in oxygen. With a view to studying further the form of the oxidation product and its connection with photoconductivity, single-crystal lead-sulphide deposits have now been prepared by condensation on rocksalt substrates. Rocksalt was chosen as likely to initiate a parallel orientation in the PbS because its axial length (5·639 A.) is within 6% of that of PbS (5·929 A.). The structure of these PbS deposits as shown by electron diffraction is described below. The results afford a picture of the mode of growth of a semiconductor condensed from its vapour on to an ionic substrate and provide a useful comparison with the growth of similarly prepared deposits of metals (Kirchner 1932, Lassen 1934, Lassen and Brück 1935, Rüdiger 1937, Wilman 1939).

§2. EXPERIMENTAL

Fresh rocksalt {001} faces about 8×5 mm. were prepared immediately before use by cleavage from an almost perfect natural crystal. The specimen was handled only with tweezers, and any stray fragments were lightly brushed off the face with a camel hair brush. Other faces used were ground on large cleavage blocks of rocksalt and smoothed on 0000 emery paper before lightly etching during a second or two in a fast stream of tap-water, followed by plunging immediately into absolute alcohol to stop further etching and to remove water and solution. Diffraction patterns of Kikuchi lines and spots elongated perpendicular to the shadow edge from such surfaces showed that the resulting smooth surfaces were free from random recrystallized NaCl.

The lead sulphide was prepared by precipitation from concentrated lead-nitrate solution by H_2S gas, followed by repeated washing with distilled water and drying in air at about 80° C. After driving off excess sulphur by heating *in vacuo*, a few milligrams were transferred to another pyrex tube and sublimed in a vacuum of the order of 10^{-3} mm. Hg on to rocksalt substrates.

The deposits were normally allowed to cool for at least five minutes before letting in air and transferring to a Finch-type electron-diffraction camera (Finch and Wilman 1937). A camera length of about 47 cm. and 50–65 kv. electrons were used. For transmission examination the rocksalt substrate was dissolved away in distilled water and the floating PbS film was picked up on a nickel gauze and dried *in vacuo* in the diffraction camera.

The rate of deposition and thickness of deposits were controlled roughly by eye to standard conditions, namely a rate of the order of 40 A./sec. and thicknesses about 50, 250 or 1000 A. (corresponding respectively to light yellow, translucent brown and opaque, or nearly so, by transmitted light). The thicknesses were estimated for a few cases by subliming known weights of PbS on to a measured area and visual comparisons with these were made for the remaining specimens.

§3. THE STRUCTURE OF THE DEPOSITS

The results are outlined in tables 1 and 2. The deposits prepared were mostly of medium thickness in view of the purpose for which they were later intended.

Deposits on apparently good ground and etched surfaces sometimes consisted of randomly disposed PbS crystals, even at the higher temperature range where strong orientation of the PbS was usually observed, all other conditions being kept constant. These results are not recorded in table 2 because they were attributed to the accidental presence of submicroscopic randomly disposed NaCl crystals on the main rocksalt substrate.

(i) *Orientations of the PbS*

In addition to the orientations listed in the tables, small proportions of randomly disposed PbS crystals were occasionally indicated by faint PbS rings in the diffraction pattern as in figure 2, but the orientation was usually very strong as shown by well-defined diffraction spot patterns, figures 1 and 6. In the case of deposits on {110} and {111} NaCl faces below about 150° c. the only orientation observed was with {111} planes parallel to the substrate with otherwise random azimuthal distribution.

(ii) *Twinning*

About one-third of the deposits on the {001} NaCl faces were found to possess strongly twinned structures on either {111} or {332} planes as shown by figure 2 and figures 3 and 4 respectively. The {332} twinning appears not to have been observed hitherto in the mineral galena. Deposits condensed from the vapour must of course grow from many nuclei formed initially on the substrate, not necessarily all in the same orientation; thus it is not clear whether these PbS crystals first grew in an orientation parallel to that of the NaCl and then twinned, or whether they were separate crystals in twin orientation. In the above cases, the crystals in the twin orientations to those which are parallel to the rocksalt are believed to arise actually by twinning from initially formed PbS crystals whose axes are parallel to those of the NaCl, because these latter crystals have in these cases developed faces (see (iii) below) which are parallel to the twinning plane. The planes in the {111} and {332} twin crystals, which would be junction planes with the NaCl {001} substrate, have lattice-point distributions which are only similar in spacing to the NaCl along one direction, i.e. the $\langle 110 \rangle$ rows of the NaCl, and growth in these orientations on the NaCl is, therefore, less likely than in the {001} orientation, though possible.

Diffraction spots due to {*hhl*} types of twinning are most prominently observed in the $\langle 110 \rangle$ azimuth, their positions being then obtained by the rotation of the main pattern (the stronger spots) about the central undeflected spot, through $2 \tan^{-1} l/h\sqrt{2}$, namely 70° 31′ for {111} twinning (figure 2), 50° 28′ for {332} twinning (figure 3). None of the seven deposits on {111} NaCl nor that on a {443} NaCl face showed any evidence of twinning, and only one on {110} NaCl showed a few extra spots not yet explained.

(iii) *Development of external crystal faces*

In most of the PbS deposits the crystals had tended to develop definite external faces to a greater or lesser extent as shown by elongation of the spots, especially those near to the shadow edge. In all cases the faces parallel to the substrate

Table 1. The structure of PbS deposits on {001} NaCl cleavage faces

Approx. temp. of substrate (° C.)	Approx. deposit thickness (A.)	No. of specimens with following characteristics									
		Total number	Orientations observed			Order of crystal diameter (A.)				Twinning observed	
						‖L ¶		⊥R¶			
			{001}†	{111}‡	Random	20–50	50–200	20–50	50–200	{111}	{332}
20	50	6	5	–	1**	2+2**	2	2+2**	2	1	–
100–200	50	3	2§	2§	1	3	–	3	–	–	–
200–300	50	3	2+1§	1§	–	3	–	1	2	1	–
	250	11	10+1§	1§	–	4	7	2	9	1	4
	1000	3	3	–	–	1	2	–	3	3	–
	250*	3*	3	–	–	–	3	–	3	–	–
	1000*	3*	3	–	–	–	3	–	3	–	–
Approx. 400	250	2	2§	2§	–	–	2	–	2	–	–

* These deposits were kept at the deposition temperature for a further 30 sec. after conclusion of deposition before being allowed to cool in the normal manner.

† Orientation with (001) planes parallel to {001} cleavage surface of the NaCl with the cubic axes parallel to those of NaCl.

‡ This denotes here {111} PbS parallel to {001} NaCl cleavage face and ⟨110⟩ lattice row parallel to ⟨110⟩ of NaCl.

§ These specimens showed both orientations together, figure 5.

¶ ‖L and ⊥R are reckoned with respect to the substrate surface.

** These specimens had a mean crystal diameter < 20 A.

Table 2. The structure of PbS deposits on {110}, {111} and {443} NaCl faces

Approx. temp. of substrate (° C.)	Approx. deposit thickness (A.)	No. of specimens with following characteristics					
		Total number	Orientations observed	Order of crystal diameters (A.)			
				‖L†		⊥R†	
				20–50	50–200	20–50	50–200
Deposits on {110} *NaCl face*							
20	50	1	Slight {111} ‖L to substrate	1	–	1	–
100–200	50	1	Random	–	1	–	1
200–300	50	1	‖L* 1	1	–	1	–
	250	2	‖L* 2	1	1	1	1
Deposits on {111} *NaCl face*							
100–200	50	1	Slight {111} ‖L substrate	1	–	1	–
	1000	1	Random	–	1	–	1
200–300	50	4	‖L* 4	4	–	4	–
	250	2	‖L* 2	2	–	2	–
Deposit on {443} *NaCl face*							
200–300	250	1	‖L* 1	1	–	1	–

* ‖L denotes PbS cubic axes parallel to those of the NaCl.
† ‖L and ⊥R are reckoned with respect to the substrate surface.

were the most prominent and were often very extensive and atomically smooth as shown by the degeneration of the spot pattern into almost continuous parallel lines perpendicular to the shadow edge. This is shown in figure 1, characteristic of those deposits which were kept at the deposition temperature (200–300° c.) for a further 30 sec. after deposition. The larger size of the crystals in these deposits was shown by the appearance of diffuse Kikuchi bands, as well as by the sharpness of the vertical streaks.

The patterns from one of the {332}-twinned deposits on {001} NaCl in the ⟨100⟩ and the ⟨110⟩ azimuths showed that the crystals in the main orientation parallel to the NaCl had developed {100}, {110}, {120}, and {335} faces, the streaks in the pattern corresponding to {335} faces making an angle of 40° with the plane of incidence in figure 3. The elongation of the twin spots perpendicular to the shadow edge showed that the twin lattices also had developed boundary faces parallel to the substrate and therefore of {335} type with respect to the cubic axes of the twin lattices. In another twinned PbS deposit having a smaller proportion of {332} twinning there were strongly developed {001} faces, and {120} faces were prominent though {110} and {335} were not observed.

In PbS deposits on {110} and {111} NaCl faces there was slight elongation of spots near and normal to the shadow edge, showing a tendency of the crystals

Figure 1. PbS on {001} NaCl; {001} orientation, <110> Az.

Figure 2. PbS on {001} NaCl; {001}, <110> Az., {111} twinning.

Figure 3. PbS on {001} NaCl; {001}, <100> Az., {332} twinning

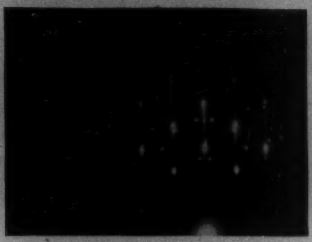


Figure 4. PbS on {001} NaCl; {001}, <110> Az., {332} twinning.

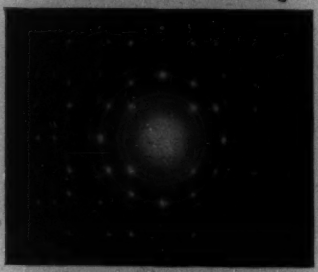


Figure 5. PbS grown on {001} NaCl. {001} and {111} orientations.

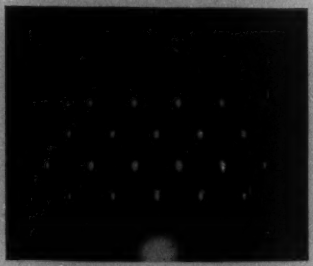


Figure 6. PbS on {110} NaCl; parallel orientation.

Figure 7. Random PbS on {111} NaCl.

Figure 8. Deposit of Figure 7 heated in vacuum to ~ 250°c.

To face page 169

in the deposit to form faces parallel to the substrate. In two specimens on {110} NaCl faces there was pronounced development of {001} boundary faces, inclined at 45° to the substrate, and the large extent of the spot pattern is also consistent with the development of such small projections on the deposit surface.

The thin PbS deposit prepared on a {443} NaCl face had all its crystals very strongly oriented with their cubic axes parallel to those of the NaCl. The spot patterns were as free from arcing as those of figures 3 and 4, though without any elongations representing definite boundary faces. The pattern previously obtained from the {443} NaCl face showed Kikuchi bands and relatively few but strong spots elongated normal to the shadow edge, showing almost perfect {443} boundary faces.

(iv) *Lattice constants*

In two cases of PbS deposits at 200–300° C. on {001} NaCl faces, the lattice constants of the PbS were measured from the patterns obtained by transmission after dissolving away the NaCl. In one case the reference material was graphite ($d_{110}=1{\cdot}230$ A.) giving $a_{PbS}=5{\cdot}931$ A., and in the other case NaCl crystallized from solution ($d_{200}=2{\cdot}820$ A.) giving $a_{PbS}=5{\cdot}927$ A. These values are close to the average of $a=5{\cdot}929$ A. previously found for PbS layers formed by passing H_2S gas over lead-nitrate or lead-acetate solution or sublimation of such material *in vacuo* in pyrex cells (Wilman 1948).

§ 4. THE EFFECT OF HEATING PbS DEPOSITS ON NaCl SUBSTRATES *IN VACUO*

It has been mentioned above that the continued heating of PbS deposits *in vacuo* at 200–300° C. after deposition on the {001} NaCl faces, only resulted in elimination of twinned structure, increase of crystal diameter and strong development of faces parallel to the substrate. No change of the main orientation was observed. It seemed possible, however, that (i) deposits consisting of random crystals as a result of low mobility of the deposit atoms on the substrate at the temperature of deposition might develop a two-degree orientation relative to the substrate lattice when heated *in vacuo*; and (ii) that the rate and kind of orientation developed should depend on the temperature at which the heat treatment is carried out and on the distribution of atoms or ions in the substrate surface.

To test this a 1000 A. thick PbS layer was deposited at 100–150° C. on a {111} NaCl face and was found to consist of random crystals about 120 A. in diameter. When this was heated for six minutes at 200–300° C. *in vacuo* it developed strong single-crystal structure (figure 8) with a {111} plane parallel to the {111} substrate surface and the cubic axes parallel to those of the NaCl. A further confirmation was obtained in the case of a 50 A. thick random layer deposited at room temperature on a {001} NaCl cleavage face. Though initially composed of random crystals of only 10–20 A. mean diameter, on heating to 200–300° C. *in vacuo* for four minutes it developed strong two-degree {001} and {111} orientation similar to that in layers deposited at this temperature on {001} NaCl. In both cases there was a large and rapid increase in crystal size at this temperature of heating, to a mean diameter of at least 300 A. though less than 500 A.

§ 5. ORIENTATION OF NaCl ON PbS

Several cases were observed, e.g. figure 6, where traces of NaCl had evidently been present on the walls of the pyrex tube and has sublimed on to the already

deposited and oriented PbS. These NaCl deposits were found in all cases to have the same orientation as the underlying PbS deposit, whether this was oriented or random.

§ 6. DISCUSSION

(i) *The orientations observed*

The orientations of the PbS crystals relative to the {001}, {110}, {111} and {443} NaCl substrates agree with the conditions generally observed, that in such cases of strongly oriented overgrowths there is a similarity of periodicity in at least one set of parallel lattice rows in the deposit and substrate crystal. The development of {111} orientations on {001} NaCl at temperatures above about 250° C. as well as the first-developed {001} orientations is analogous to the case of silver deposits formed on {001} NaCl faces over a similar range of temperatures (Brück 1936). On {111} and {443} NaCl faces, the consistent growth of PbS with cubic axes parallel to those of the NaCl, and not in more than one orientation geometrically equivalent relative to the surface plane, is analogous to the continued growth of NaCl on NaCl, building up a continuous pattern of positive and negative ions. Thus it would seem that not only the surface atoms govern the orientation in the growth of such ionic crystals but also those in the underlying net-plane.

(ii) *Twinning*

The proportion of specimens having a {111} or {332} twinned structure is only about 30% compared with almost 100% for those of silver. This may perhaps be associated with the smaller difference of lattice constants of the substrate and PbS deposit (cf. Wilman 1939). On the other hand, it seems likely that here the PbS crystals grow initially in an orientation parallel to the NaCl and twin during their further growth, whereas silver is strongly {111} twinned in very thin layers; but the crystals in the four twin orientations grow together (Menzer 1938), so that surface regions of thick silver layers are formed of crystals whose axes are parallel to those of the NaCl.

(iii) *Lattice constants*

The present estimations from transmission patterns show more accurately than was previously possible from reflection patterns (Wilman 1948) the striking constancy of the lattice axial length, 5·929 A., of PbS prepared both by chemical deposition and by sublimation *in vacuo*.

(iv) *Interpretation of deposit structure in terms of a hypothesis of lowest potential energy state modified by surface mobility and rate of deposition of atoms from the vapour stream*

It is now clear that oriented deposits must arise through the initially deposited atoms taking up stable positions where they have lowest potential energy on substrate crystal surfaces which consist of a periodic distribution of atoms or ions. Similar views have been expressed for example by Stranski (1928). In like manner further atoms take up packing positions having least potential energy, and thus consistent with the normal deposit crystal structure and atomic distribution, i.e. orientation, in the first layer. The attainment of such a state is, however, dependent especially upon the rate of deposition and on the thermal movements and the strength of forces between the deposit and substrate ions compared with those of the deposit and substrate alone. Increase in temperature

of the substrate, by increasing the amplitude of thermal vibrations of the substrate surface atoms, should assist the growth of strongly oriented deposits, and this has indeed been confirmed by many experiments including the present results on PbS.

The effect of rate of deposition on the number and size of the deposit crystals at any stage of the deposition on single crystal or amorphous substrates has not been quantitatively explored in the present experiments on PbS, but the general observations so far made may be explained as follows: The atoms first falling on the substrate have a certain degree of mobility over the surface before losing most of the kinetic energy which they possessed on arrival; thus, sooner or later, they meet other atoms and aggregate to form small nuclei. The greater the number of atoms incident on unit surface area per unit time, the larger will be the number of these nuclei formed per cm^2/sec. and the smaller their lateral extent, and also the smaller the distances between these nuclei on the still uncovered substrate. The higher the rate of arrival of the atoms per cm^2/sec., the greater is the number of atoms bombarding each nucleus per sec., and these will therefore hinder the development of the orientation which would be imposed by the substrate and result in a correspondingly weaker orientation or even a completely random crystal deposition. This picture is analogous to the Brownian movement of small particles in a liquid due to bombardment by the atoms of the liquid surrounding them.

The above experiments on PbS show that, even with a relatively high rate of deposition of about 40 A. thickness per second, strongly oriented PbS deposits on NaCl were obtained at substrate temperatures of about 150° C. As in the case of silver and other metals under similar conditions, random or weakly oriented PbS deposits tend to be formed at or near room temperature; and no strong development of orientation with increase of thickness was observed, such as occurs when the crystals develop large plane faces perpendicular to the incident beam direction. (Kirchner 1932, Burgers and Dippel 1934, Beeching 1936, Burgers and van Amstel 1936). These results, together with the observation that random-crystalline PbS deposits do not develop orientation unless heated well above room temperature, suggest that the PbS molecules have low mobility over NaCl surfaces at room temperature, but that the mobility rapidly increases with temperature.

The PbS deposits prepared at room temperature tended to be orientated apparently because the substrate surface temperature became more or less raised by absorption of radiant energy from the hot test-tube wall from which the PbS was being sublimed. The temperature of the substrate at its surface is difficult either to determine or control. It is a result of equilibrium between energy falling on to the surface by radiation from the oven—or filament—source of atoms, and energy leaving the surface by re-radiation and by conduction to the lower layers of the substrate. It was found that PbS deposits about 50 A. thick on an NaCl-crystal substrate 2 to 3 mm. thick were often weakly oriented, but when 10 mm. thick NaCl crystals of about the same area were used, the 50 A. thick PbS deposits obtained were of quite randomly disposed crystals only 10–20 A. in diameter. This demonstrates the need for using substrates of high thermal capacity, if necessary backed by a metal mass of high conductivity, to ensure that the substrate surface temperature is not much higher than that of the rest of the substrate.

(v) *Recrystallization of initially random PbS deposits*

According to the above view of the dependence of crystal growth on the deposit conditions, we would expect that even if the crystals of the deposit are initially randomly disposed, subsequent heat treatment should enable rearrangement of the atoms to take place in such a way that the state of least potential energy with the substrate, consistent with the surface mobility, is attained. This state should be the same as that which would have been obtained if the layer had been deposited at this temperature. The results of the experiments in §4 do, in fact, lend further support to the hypothesis that the lowest potential energy state tends to be attained as far as the limited surface mobility of the deposit atoms permits. In this connection it is noteworthy that both {001} and {111} orientations of the PbS were developed together on the {001} NaCl face when heated *in vacuo* to 200–300° C.

The fact that random crystalline (or amorphous) deposits on heating take up an orientation characteristic of the substrate and its surface temperature, which under these conditions becomes that of the heating enclosure, enables the relation to be studied between the orientation of thin deposits and the nature and surface temperature of the substrate, without the uncertainty of measurement of the actual surface temperature during the deposition process. This should be especially helpful in the case of deposit materials which show a variety of orientations within a relatively narrow temperature range.

§7. SUMMARY

The structure of PbS deposits condensed from the vapour *in vacuo* on to {001}, {110}, {111} and {443} rocksalt faces has been investigated by electron diffraction. The results obtained agree with the theory that the deposit atoms take up positions of least potential energy relative to the substrate, as far as is permitted by the disturbing effects of collisions of incident atoms with the initial deposit crystal nuclei and by the limited surface mobility of the deposited atoms over the substrate. Further confirmation of our theory is afforded by changes of orientation of initially random deposits when heated *in vacuo*, and lead to the conclusion that such experiments should enable a more precise investigation to be made of the relation between the deposit orientation and the temperature of the substrate surface.

ACKNOWLEDGMENTS

This work forms part of a study of the fundamental mechanism of crystal growth being carried out in this Laboratory and it is a pleasure to thank Professor G. I. Finch for his interest and encouragement.

A. J. Elleman also thanks the Department of Scientific and Industrial Research for a grant which enabled him to devote his full time to this research.

REFERENCES

Beeching, R., 1936, *Phil. Mag.*, **22**, 938.
Brück, L., 1936, *Ann. Phys. Lpz.*, **26**, 233.
Burgers, W. G., and van Amstel, J. J. A., 1936, *Physica*, **3**, 1057.
Burgers, W. G., and Dippel, D. J., 1934, *Physica*, **1**, 549.
Finch, G. I., and Wilman, H., 1937, *Ergebn. exakt. Naturw.*, **16**, 353.
Kirchner, F., 1932, *Z. Phys.*, **76**, 9, 576.
Lassen, H., 1934, *Phys. Z.*, **35**, 172.
Lassen, H., and Brück, L., 1935, *Ann. Phys., Lpz.*, **22**, 65.
Lee, E., and Parker, R. C., 1946, *Nature, Lond.*, **158**, 518.

MENZER, G., 1938, *Naturwissenschaften*, **26**, 385 ; *Z. Kristallogr.*, **99**, 378 and 410.
RÜDIGER, O., 1937, *Ann. Phys., Lpz.*, **30**, 505.
SOSNOWSKI, L., SOOLE, B. W., and STARKIEWICZ, J., 1947, *Nature, Lond.*, **160**, 471.
SOSNOWSKI, L., STARKIEWICZ, J., and SIMPSON, O., 1947, *Nature, Lond.*, **159**, 818.
STARKIEWICZ, J., SOSNOWSKI, L., and SIMPSON, O., 1946, *Nature, Lond.*, **158**, 28.
STRANSKI, I., 1928, *Z. phys. Chem.*, **136**, 259.
WILMAN, H., 1939, *Proc. Phys. Soc.*, **51**, 625 ; 1948, *Ibid.*, **60**, 341.

A Determination of the Ratio of the Masses of π- and μ-Mesons by the Method of Grain-counting *

BY C. M. G. LATTES, G. P. S. OCCHIALINI AND C. F. POWELL
H. H. Wills Physical Laboratory, University of Bristol

MS. received 10 *April* 1948

§ 1. INTRODUCTION

IN a recent article (Lattes 1947), we gave an account of a determination, by the method of grain-counting, of the ratio of the masses of the π- and μ-mesons observed in photographic emulsions. We have found more than thirty events in which the secondary μ-meson produced by a π-meson at rest comes to the end of its range in our emulsions. This new material, of which the main features are summarized in table 1, allows us to increase the accuracy of the determination,

Table 1. Ranges of π- and μ-mesons in individual examples of the μ-decay

Event No.	Ranges of π-meson (microns)	Ranges of μ-meson (microns)	E_μ (MEV.)	Event No.	Ranges of π-meson (microns)	Ranges of μ-meson (microns)	E_μ (MEV.)
1	134	613	4·08	17	477	585	3·97
2	1287	615	4·09	18	12	685	4·36
3	124	591	4·00	19	52	644	4·21
4	106	633	4·16	20	62	594	4·00
5	453	615	4·09	21	230	570	3·91
6	105	567	3·91	22	115	570	3·91
7	906	606	4·05	23	488	680	4·34
8	248	627	4·13	24	96	667	4·29
9	231	648	4·22	25	409	626	4·13
10	81	590	3·99	26	91	623	4·12
11	57	609	4·06	27	990	602	4·04
12	634	567	3·90	28	92	619	4·10
13	48	611	4·06	29	15	598	4·01
14	1297	603	4·04	30	91	639	4·18
15	72	591	4·00	31	2290	660	4·27
16	57	590	3·99				

Mean values : $R_\mu = 614\ \mu$; $E_\mu = 4{\cdot}08$ MEV.

and in this article we describe in more detail the method of making the measurements and the results which follow from them. We find that the mass ratio $m_\pi/m_\mu = 1{\cdot}65 \pm 0{\cdot}11$, the limits corresponding to the "probable error" arising

* An account of the main features of a number of lectures given in Copenhagen in December 947, in Zurich in January 1948, and in a number of colloquia in English Universities.

from statistical fluctuations alone. It is reasonable to assume that the μ-mesons appear as a result of a spontaneous decay of the heavier π-mesons, and that the former are identical in type with the ordinary cosmic-ray mesons of mass $200\,m_e$. It then follows from the present results that the momentum balance in the decay is provided by a neutral particle, a neutral meson with a rest-mass equal to $115 \pm 30\,m_e$, where m_e is the mass of the electron. The observations also allow us to draw conclusions about the decay constants of the different particles, and we have thus determined the following limits for the half-value periods: $\tau_\pi > 0{\cdot}4 \times 10^{-8}$ sec., $\tau_\mu > 2 \times 10^{-10}$ sec., and $\tau_\nu > 2{\cdot}6 \times 10^{-21}$ sec., where τ_ν is the half-value period for the postulated neutral meson.

§ 2. PRINCIPLE OF THE METHOD

In determining the quantity m_π/m_μ by grain-counts, we make the usual assumption that a particle with a charge $|e|$, energy E Mev., and range in the emulsion R microns, loses energy at a rate which depends only on its velocity, v, and is independent of its mass, m : $\partial E/\partial R = f_1(v)$. It is then easily shown that $\partial E/\partial R = f_2(R/m)$ and $E = m\phi(R/m)$.

We assume, secondly, that if $N(R)$ is the total number of grains in the track of a particle of range R microns, the number of grains per micron, $\partial N/\partial R$, in an element of length of the trajectory, is a function only of the rate of loss of energy, $\partial E/\partial R$, of the particle. It follows that

$$N = mF(R/m). \qquad \ldots\ldots(1)$$

Equation (1) allows us, in principle, to determine the ratio m_π/m_μ from the observed numbers of grains in the track of π- and μ-mesons.

In making grain-counts, we employ a $\times 95$ "Achromatic" objective with $\times 10$ eyepieces, and count the numbers of grains in successive intervals along the trajectory of length $50\,\mu$. These intervals are measured along the projection of the track on the plane of the emulsion. With this procedure we shall tend to over-estimate the grain-density in a track which dips relative to the plane of the emulsion, through "fore-shortening". We have therefore confined our measurements to those events in which the angle of dip, β, at every point in the trajectories of the primary and secondary particles is never greater than 15°. The errors due to "fore-shortening" are then small.

A second source of error exists owing to the variation of the degree of development of an emulsion with depth which occurs when a plate is processed by conventional methods. This has the consequence that the grain density may be different in the tracks of two particles produced contemporaneously, in regions of equal ionization, if they are situated at different depths in the emulsion.

Thirdly, as we pointed out in a previous communication, there is not a completely random distribution of grains in the gelatine in the emulsions at present available. "Islands" appear to exist in which the concentration of silver-halide grains is significantly higher, or significantly lower, than we should expect to find as a result of random fluctuations alone. This feature seems to be particularly noticeable in boron-loaded plates. The measurements show, however, that the influence of this factor on the total numbers of grains in the relatively long tracks on which our observations are made is not more serious than the purely statistical fluctuations associated with the finite number of grains in a track.

All the above sources of error will effect the measurements on π- and μ-mesons indiscriminately; and the errors they produce in the determination of the mean

value of the mass-ratio will be small if the measurements are made on a sufficiently large number of events. They contribute, however, to the variation among the individual values of the mass-ratio.

The following difficulties cannot at present be avoided. First, at the end of the range of a particle the ionization is sufficiently intense to produce an almost continuous succession of grains which cannot be individually resolved. We refer to this effect as "clogging". No accurate counts can be made in this region of a track, and we therefore adopt the following convention. We define the number of grains in a segment of the trajectory of length r, in which there is a contiguous succession of grains, as equal to r/a, where a is the mean "diameter" of the grains. In meson tracks "clogging" is serious only in the last 50 μ from the end of the range, and for such short tracks no reliable mass determinations are possible.

Secondly, when the rate of loss of energy of the particle is very small, the grain density becomes so low that the counts become unreliable because of the general "background" of grains. This sometimes makes it difficult to decide whether a particular grain was produced by the passage of the particle or is due to "fog". In our emulsion this difficulty begins to become appreciable at 400 μ and 650 μ, respectively, from the ends of the ranges of μ- and π-mesons.

§3. CONSISTENCY OF GRAIN-COUNTS

Even in those parts of the trajectory of a particle in which "clogging" is not serious, the individual grains cannot always be completely resolved. It is therefore necessary to adopt a convention that a particular size of grain, for example, shall be regarded as two, that very small grains shall be neglected, etc. We have found in practice that if different observers establish a common convention, by reference to suitable photo-micrographs of characteristic tracks, it is possible to obtain the following degree of consistency between different measurements. An experienced observer can determine the total number of grains in the last 400 μ of the track of a particular meson with a consistency of 3%, and there is a similar degree of consistency between the observations of two different observers. In practice we have made a succession of five determinations of the number of grains, N_π and N_μ, in each of the tracks of the π- and μ-mesons of a particular event, making counts on the two tracks alternately. The ratio of the numbers, N_π/N_μ, on which the mass ratio depends, is then found to be consistent to about 3%. Because of the finite number of grains in a track, the statistical fluctuation in the numbers N_π or N_μ to be expected in going from one event to another is of the order of 6%, and the difference in the average number of grains in equal lengths of the tracks of π- and μ-mesons about 13%.

§4. RESULTS

In table 2 we show the results of grain counts on those events which satisfy the conditions that at no point in either of the trajectories is the angle of dip greater than 15°. The table shows the number of grains, $N(R)$, in a residual range R microns—for intervals 50 μ long—up to a maximum of 400 μ for μ-mesons and of 650 μ for π-mesons. The numbers in each vertical column correspond to the measurements on a particular event, so that the results for the two types of mesons can be compared.

It will be seen that the corresponding results for different events are subject to considerably greater fluctuations than those corresponding to statistical effects

Table 2. Grain counts on the track of π- and μ-mesons

Event No.	2	14	7	12	5	17	8	9	1	4	3
Range (microns)	No. of grains in tracks of										
	μ-mesons										
50	66	55	54	72	61	60	60	66	45	53	69
100	109	97	90	127	105	104	105	116	80	91	118
150	144	125	111	166	147	133	146	164	97	125	162
200	179	157	132	204	180	162	183	205	119	154	205
250	211	184	157	235	216	198	216	242	142	179	249
300	236	201	184	266	239	233	243	276	167	196	287
350	265	216	201	287	266	261	274	319	188	224	319
400	289	233	220	312	293	285	301	341	204	251	339
	π-mesons										
50	75	57	66	72	68	65	71	80	41	58	75
100	129	96	114	120	121	108	126	142	81	101	129
150	171	136	151	159	179	154	175	194	111	140	
200	218	172	180	198	222	196	214	235	135		
250	260	200	208	234	250	222	258				
300	298	231	233	266	278	253					
350	330	257	243	300	307	286					
400	363	282	268	332	340	313					
450	391	306	292	367	367						
500	416	329	313	400							
550	451	355	332	432							
600	478	376	343	466							
650	508	392	359								

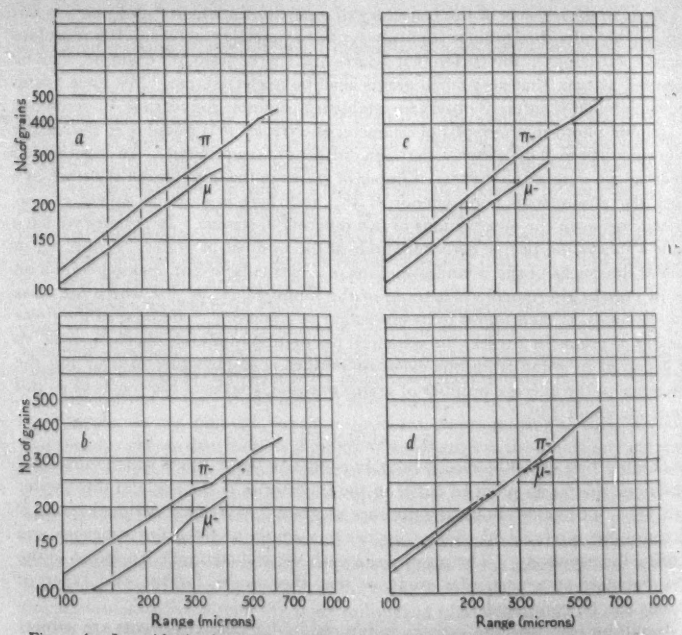


Figure 1. Logarithmic plots of the number of grains in the tracks of π- and μ-mesons for four examples of the μ-decay. (*d*) shows the results for the event which displays the least difference in the number of grains in the tracks of the associated π- and μ-mesons.

alone. We attribute these variations to "fading" and to the other factors discussed above. The results for four typical events are represented in figure 1, in which $\log N$ is plotted against $\log R$. Figure 1 (*d*) corresponds to the event which displays the least difference between the results for the π- and μ-mesons.

In order to compare the results for the different events, we normalize the values for the different observations. We first determine a normalizing factor for each event, k_i, defined as the ratio of the average number of grains in the last 400 μ of the different μ-mesons, $N_\mu^m(400)$, to the number of grains, $N_\mu^i(400)$, in the same length of the particular μ-meson under consideration: $k_i = N_\mu^m(400)/N_\mu^i(400)$. Table 3 shows the values $\mathcal{N}(R)$ obtained by multiplying the numbers $N(R)$ for

Table 3. Values of $\mathcal{N}(R)$, for π- and μ-mesons, after normalizing to correspond to a constant value of $\mathcal{N}_\mu(400)$

Event No.	2	14	7	12	5	17	8	9	1	4	3	$\mathcal{N}^m(R)$
k_i	0·956	1·197	1·268	0·894	0·952	0·979	0·927	0·818	1·368	1·112	0·823	
Range (microns)	μ-mesons											
50	64	66	69	64	58	59	56	54	62	59	57	60±2
100	105	116	114	114	100	102	97	95	110	101	97	104±2
150	139	150	141	148	140	130	135	134	133	139	133	138±2
200	173	188	167	182	171	159	170	168	163	171	169	171±3
250	204	219	199	210	206	194	200	198	194	199	205	202±3
300	228	241	233	238	228	228	225	226	228	218	236	230±3
350	256	259	255	257	253	256	254	261	257	249	263	256±3
400	279	279	279	279	279	279	279	279	279	279	279	279±3
	π-mesons											
50	72	68	84	64	65	64	66	65	56	64	62	65±2
100	124	115	145	107	115	106	117	116	111	113	106	116±2
150	165	163	191	148	170	151	162	159	152	156		160±3
200	210	206	228	177	211	192	198	192	185			198±3
250	251	239	264	209	238	217	239					232±4
300	288	277	295	238	265	248						263±4
350	318	308	308	268	292	280						290±5
400	350	338	340	297	324	306						320±5
450	377	366	370	328	349							348±6
500	401	394	397	358								375±6
550	435	425	421	386								404±7
600	461	450	435	417								426±7
650	490	469	455									449±8

The values of $\mathcal{N}_\pi^m(R)$ for π-mesons have been obtained by determining the average number of grains, $\Delta\mathcal{N}_\pi^m(R)$, in successive intervals of length 50 μ, and summing over the total range R.

the different events given in table 2, by the appropriate normalizing factor k_i, i.e. $\mathcal{N}(R) = k_i N(R)$. It may be emphasized that the mass ratio m_π/m_μ depends only on N_π/N_μ, and the values deduced from the observations on the individual events will not be changed by the use of the normalizing factor. Table 3 also contains the mean values for the number of grains, $\mathcal{N}_\pi^m(R)$, in a range R microns. The results are represented diagrammatically in figure 2, in which values of $\log \mathcal{N}^m(R)$ are plotted against $\log R$, for both π-mesons (upper line) and μ-mesons (lower line).

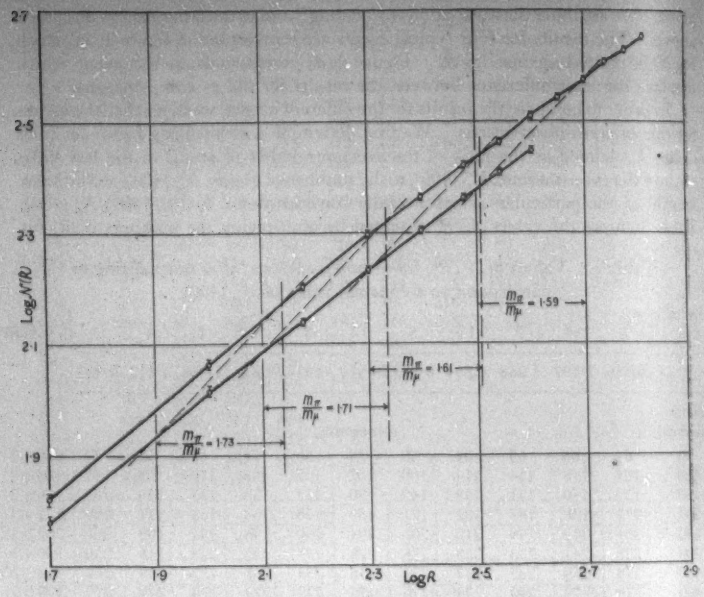


Figure 2. Plots of $\log \mathcal{N}_\pi(R)$ and $\log \mathcal{N}_\mu(R)$ against $\log R$. The upper and lower curves are drawn through the results for π- and μ-mesons respectively.

§ 5. DETERMINATION OF m_π/m_μ

The determination of the quantity m_π/m_μ can be carried out by any one of the following methods, the first of which was suggested to us by D. H. Perkins. It may be emphasized that they represent merely alternative modes of analysis, and not independent determinations.

(*a*) *Perkins' method*

It follows from the equation $N = mF(R/m)$ that if we establish a graph showing the relationship between $\mathcal{N}_\mu^m(R)$ and R for μ-mesons, we can deduce the corresponding curve for π-mesons for any assumed value of the mass-ratio m_π/m_μ. For each point $(\mathcal{N}_\mu^m(R), R)$ in the graph for μ-mesons we construct the corresponding point $\mathcal{N}_\mu^m(R)m_\pi/m_\mu$, $Rm_\pi/m_\mu)$. The locus of these points then corresponds to the relation between $\mathcal{N}_\pi^m(R)$ and R for π-mesons. Alternatively, if we plot the relation between $\log \mathcal{N}_\mu^m(R)$ and $\log R$, the corresponding values of $\log \mathcal{N}_\pi^m(R)$ and $\log R$ are given by a simple translation of magnitude $\log(m_\pi/m_\mu)$ in both coordinates.

Conversely, if the relations between $(\log \mathcal{N}_\pi^m(R), \log R)$ and $(\log \mathcal{N}_\mu^m(R), \log R)$ have been established experimentally, values of m_π/m_μ can be deduced by drawing a succession of lines at 45° to the axes and measuring the translations, AB in figure 2, to which they correspond. The values so obtained are shown in table 4.

Table 4. Values of m_π/m_μ deduced from normalized grain-counts

Range (microns)	50	100	150	200	250	300	350	400
m_π/m_μ	1·70	1·67	1·73	1·63	1·58	1·55	1·60	1·64
"Probable error" (±)	0·18	0·14	0·13	0·13	0·13	0·12	0·12	0·11

The various values are not independent, and that with the greatest statistical weight is $1{\cdot}64 \pm 0{\cdot}11$.

The values of m_π/m_μ for the different values of R_μ are not independent, and that with greatest statistical weight is $m_\pi/m_\mu = 1{\cdot}64 \pm 0{\cdot}12$, the limits corresponding to the magnitude of the "probable error".

(*b*) *Determination of the form of the function* $N = mF(R/m)$

It will be seen from figure 2 that the relationships between $\log \mathcal{N}_\mu^m(R)$ and $\log R$ and between $\log \mathcal{N}_\pi^m(R)$ and $\log R$ are linear for values of R greater than $100\,\mu$ to within the standard deviations of the measurements. We can therefore write

$$\mathcal{N} = km(R/m)^n \quad \text{or} \quad \mathcal{N} = km^{1-n}R^n. \qquad \ldots\ldots(2)$$

We have determined the value of n from the best straight line drawn through the points representing the measurements on π- and μ-mesons in figure 3, and have obtained the values $n_\pi = 0{\cdot}701 \pm 0{\cdot}014$, $n_\mu = 0{\cdot}721 \pm 0{\cdot}014$, and the mean value $n = 0{\cdot}711 \pm 0{\cdot}011$. Equation (2) can therefore be written $\mathcal{N}^m = km^{0{\cdot}289}R^{0{\cdot}711}$. From the position of the straight lines in figure 2 we find that $km_\mu^{0{\cdot}289} = 3{\cdot}95 \pm 0{\cdot}05$ and $km_\pi^{0{\cdot}289} = 4{\cdot}53 \pm 0{\cdot}07$, so that $\mathcal{N}_\pi^m = 4{\cdot}53R^{0{\cdot}711}$ and $\mathcal{N}_\mu^m = 3{\cdot}95R^{0{\cdot}711}$. The values of $\mathcal{N}_\pi^m$ and $\mathcal{N}_\mu^m$ calculated from this formula for different values of R show no significant difference from those based on the grain-counts given in table 3, except for values of R less than $100\,\mu$.

Alternatively we can plot the quantity $a = \mathcal{N}^m/R^n$ from the observed values of $\mathcal{N}^m$; a should be a constant for the results for a given type of particle

$$a = \mathcal{N}^m/R^n = km^{1-n}. \qquad \ldots\ldots(3)$$

The values obtained are shown in table 5 and in figure 3 for both π- and μ-mesons.

Table 5

Range (microns)	$a_\mu = \dfrac{\mathcal{N}_\mu^m(R)}{R^{0{\cdot}711}}$	$a_\pi = \dfrac{\mathcal{N}_\mu^m(R)}{R^{0{\cdot}711}}$
50	3·73 ± 0·10	4·10 ± 0·10
100	3·94 ± 0·08	4·39 ± 0·08
150	3·92 ± 0·07	4·55 ± 0·08
200	3·95 ± 0·06	4·57 ± 0·08
250	3·98 ± 0·06	4·58 ± 0·08
300	3·95 ± 0·05	4·58 ± 0·08
350	3·97 ± 0·05	4·50 ± 0·08
400	3·94 ± 0·05	4·52 ± 0·08
450		4·52 ± 0·08
500		4·53 ± 0·08
550		4·55 ± 0·08
600		4·51 ± 0·08
650		4·49 ± 0·08
Weighted means :	$a_\mu = 3{\cdot}95 \pm 0{\cdot}05$;	$a_\pi = 4{\cdot}53 \pm 0{\cdot}08$.

Table 6

Individual values of the ratio m_π/m_μ

Event No.	m_π/m_μ	Error (%)	Normalized "weight"
2	2·06 ± 0·36	17·4	1·24
14	1·52 ± 0·27	17·6	1·21
8	2·04 ± 0·45	17·7	1·19
12	1·07 ± 0·19	18·0	1·15
5	1·75 ± 0·33	18·7	1·07
17	1·45 ± 0·28	19·3	1·00
8	1·84 ± 0·38	20·5	0·89
9	1·65 ± 0·36	21·8	0·79
1	1·37 ± 0·30	22·0	0·77
4	1·82 ± 0·42	23·3	0·69

Weighted mean : $1{\cdot}66 \pm 0{\cdot}11$.

In this case, also, the different values are not independent, so that the "probable error" is not reduced by taking the mean.

From equation (3) we can write

$$a_\pi/a_\mu = [m_\pi/m_\mu]^{0\cdot289} = 1\cdot145 \pm 0\cdot024; \qquad m_\pi/m_\mu = 1\cdot62 \pm 0\cdot12.$$

We can calculate values of $\mathcal{N}_\pi^m$, corresponding to the observed values of $\mathcal{N}_\mu^m$, from the formula $N = km^{0\cdot289}R^{0\cdot711}$, assuming that $m_\pi/m_\mu = 2\cdot1$, $1\cdot65$, $1\cdot32$ in turn. The corresponding values of the quantity $\mathcal{N}_\pi^m/R^n$ are represented by the dotted lines in figure 3 : (*a*) corresponds to the case in which the π-meson disintegrates to produce a neutral meson, equal in mass to the μ-meson, while (*c*) corresponds to the assumption that the momentum balance in the μ-decay is provided by an emitted photon.

§6. INDIVIDUAL VALUES OF THE RATIO m_π/m_μ

By using the equation $N = km^{0\cdot289}R^{0\cdot711}$ we can determine a value of m_π/m_μ from the observations on the individual examples of the μ-decay, and the values thus obtained are summarized in table 6. The results are also represented in figure 4. In this diagram the size of the oblong, corresponding to a particular determination, has an area proportional to its statistical weight. It will be seen that of the ten individual determinations, only one has a magnitude less than $1\cdot32$, and not one greater than $2\cdot1$.

§7. THE EXISTENCE OF A NEUTRAL MESON

The three methods which have been employed in the analysis of our observations are consistent in that they show no significant differences from the mean value $1\cdot64 \pm 0\cdot11$. Although the fluctuations of the individual values are so large that we cannot regard our results as final, they strongly suggest that the momentum balance in the process of μ-decay is provided by a neutral particle and not by a photon, and that this particle is less massive than the μ-meson. The application of the conservation laws of energy and momentum allows us to calculate the mass, m_ν, of this "neutretto", and we thus obtain the value $m_\nu = 115 \pm 30\, m_e$.

Recent photographs by Anderson, Adams, Lloyd and Rau (1947) show that the cosmic-ray mesons brought to rest in the gas of an expansion chamber decay with the emission of an electron of energy 25 MeV. Assuming the mass of these mesons to be $200\, m_e$, this result is consistent with the assumption that a neutral particle with a mass of about $130\, m_e$ is ejected in the opposite direction to the electron in the decay process. The agreement, within the limits of experimental error of the masses of the neutral particles indicated by the two experiments, is suggestive. We have therefore considered the consequences of assuming that the two particles are of the same type, and that the ordinary cosmic-ray mesons are identical with our μ-mesons. The decay sequence that follows from these assumptions is represented in figure 5. Since the electron has half-integral spin we can conclude from the process $\mu \rightarrow e + \nu$ that μ- and ν-mesons differ by half-integral spin. It then follows, from the process $\pi \rightarrow \mu + \nu$, that π-mesons have half-integral spin.

§8. THE DECAY CONSTANTS OF MESONS

(*a*) *π-mesons*

The distribution in range of the μ-mesons, of which details are given in table 1, are represented in figure 6(*a*). Following Livingston and Bethe, we can define a straggling coefficient, s, by the relation $s = \sqrt{(\pi/2)}\sqrt{\{\Sigma_n(R_l - R_m)^2/n\}}$, where R_l is the range of the lth μ-meson, the observed mean value of all the n μ-mesons

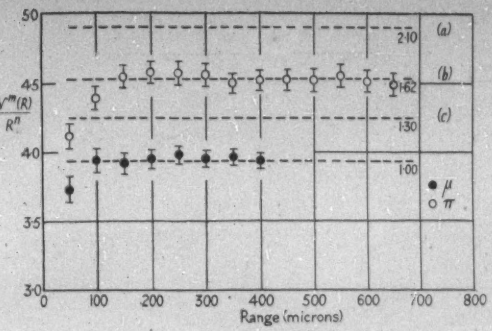


Figure 3. $n=0{\cdot}711\pm0{\cdot}010$; $N(R)=km^{1-n}R^n$.

Figure 5.

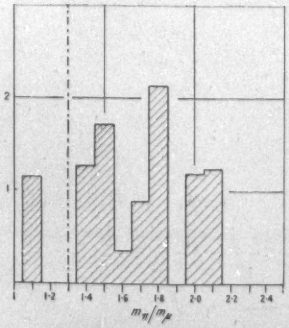


Figure 4. Individual values of mass ratio, m_π/m_μ, as determined by grain-counts.

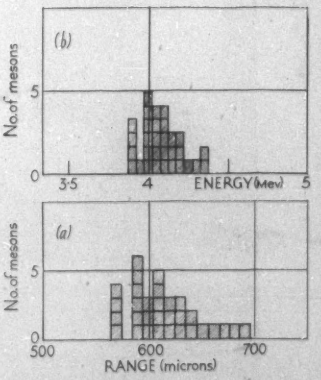


Figure 6. (*a*) Distribution in range of μ-mesons arising from the decay of π-mesons at rest. (*b*) Distribution transformed to an energy scale, using extrapolated range–energy relation for protons.

To face page 180

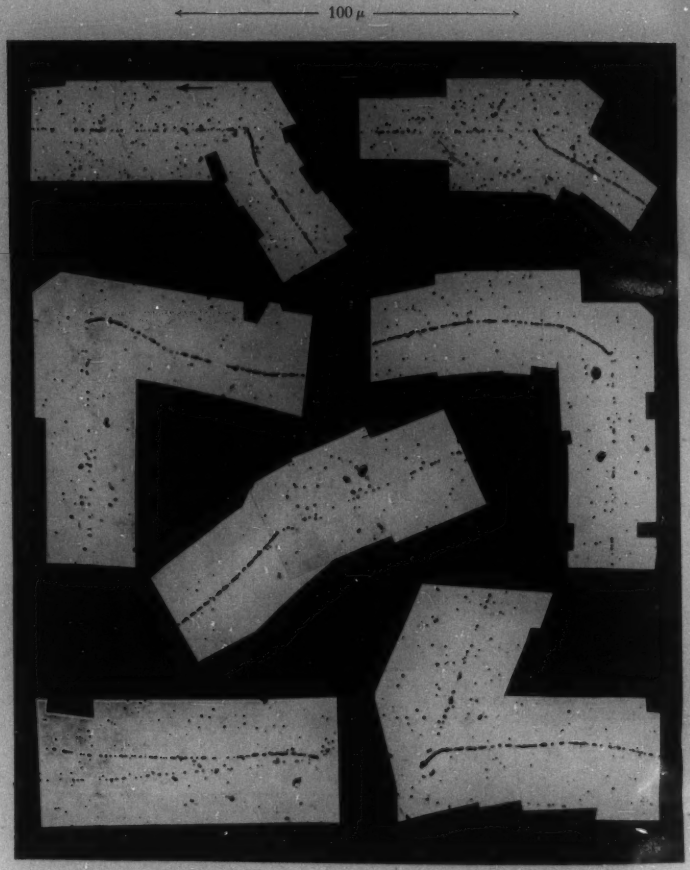


Figure 7. Seven examples of photo-micrographs of the decay of π-mesons, showing that the direction of motion of the μ-meson passes within 1μ of the last grain in the track of the π-meson.

being R_m. The measurements shown in table 1 are thus found to correspond to a value of s given by $s/R_m = 6{\cdot}24\%$.

Only extrapolated values of the range–energy relation are available, for particles of high velocity and charge $|e|$, in the "Nuclear Research" emulsions. These are given with sufficient accuracy by the equation $E = 0{\cdot}262 M^{0{\cdot}425} R^{0{\cdot}575}$, where E is the energy in MEV. and R the range in microns of a particle of M mass-units (proton = 1). It follows from this equation that $\delta E/E = 0{\cdot}575 \delta R/R$. Writing $\delta R/R = s/R_m = 0{\cdot}0624$ we find, that $\delta E/E = 0{\cdot}0359$. This value does not differ significantly from that corresponding to the width at half-maxima of the distribution in energy of the μ-mesons shown in figure 6(*b*), which has been calculated from the observed range distribution on the assumption that the mass of the μ-mesons equals $200\, m_e$. The new observations therefore confirm the previous conclusions that the μ-mesons are emitted with values of the velocity which are constant within narrow limits, for the observed straggling coefficient is of the same order of magnitude as that suggested by the classical approach to the problem.

Now the directions of ejection of the μ-mesons are oriented at random with respect to the line of motion of the μ-mesons at the extreme end of their range. Let the difference in the direction of motion of the two particles, at the point of juncture of the two tracks, be θ. If the π-mesons come to "rest" before decaying, this angle is without significance, for the particles will suffer frequent changes in direction, due to scattering, when moving at low velocities at the end of the range. If, however, a π-meson decays in flight, we should expect the velocity of ejection of the μ-meson, relative to the emulsion, to depend on the value of θ. In these circumstances, the range of the ejected mesons would be very sensitive to the direction of emission relative to the line of motion of the parent π-meson.

We have calculated the magnitude of the effect to be anticipated, making the simplifying assumption that the π-mesons always decay when 5 μ from the end of their range. It is easily shown that in these circumstances the distribution in range of the μ-mesons would extend from about 200 μ to 1200 μ. Further, we have examined the distribution in range of the μ-mesons in these events for which $\theta < \pi/2$, and, separately, those for which $\theta > \pi/2$. It is found that there is no significant difference between the two distributions. We therefore conclude that the π-mesons, in the observed events, were at or very near the end of their range when they decayed.

We cannot conclude, however, that the observed events represent all the π-mesons which, having decayed in the emulsion, led to the production of a μ-meson which stopped in the emulsion. If, for example, a π-meson decayed when its residual kinetic energy exceeded 2 MEV., it would be difficult to distinguish the event from one in which a meson is scattered by a nuclear collision. We believe, however, that if the residual range of the π-meson were less than 100 μ at the instant of decay, we should recognize the event because of the sudden decrease of the grain density in the track immediately after the point of "scattering". We therefore conclude that in twenty cases a π-meson traversed a distance of 100 μ at the end of its range without decaying, and that in no case did it decay in this interval. The time of flight of such a meson is 3×10^{-12} sec., and we can therefore write $20 \times 3 \times 10^{-12} \times k_\pi < 1$, where k_π is the decay constant of the π-mesons. It follows that $k_\pi < 1{\cdot}6 \times 10^{10}$ and $\tau_\pi > 0{\cdot}4 \times 10^{-10}$ sec.

In every case in which the μ-meson reaches the end of its range in the emulsion it stops without leading to a disintegration with the emission of heavy particles. This suggests that the π- and μ-particles are positively charged—for the Coulomb repulsion would then prevent their close approach to a nucleus when moving with low velocities, and their interaction with it—and we shall assume this to be the case.*

We have now observed 150 events in which a π-meson, at the end of its range, produces a μ-meson. In every case we can distinguish no displacement of the line of motion of the ejected μ-meson from the last grain in the track of the π-meson (see figure 7). We regard this as evidence that in no case did the π-meson, at the end of its range, diffuse a distance greater than 2μ before decaying. If it is assumed that a positively charged positive meson diffuses through the solid material of the emulsions as through a heavy gas, an estimate can be made of an upper limit to the half-period value τ_π. The result so obtained is $\tau_\pi > 5 \times 10^{-7}$ sec. Dr. Frohlich has pointed out to us, however, that the problem of the motion of a positively charged meson in a solid substance is one of great complexity. It is likely that a positive meson will be "trapped" when brought to rest in a solid and that its diffusion will thus be prevented. We can therefore attach little weight to the limit obtained by this method.

(*b*) *μ-mesons*

We have seen that in thirty cases the μ-mesons reach the end of their range in the emulsion. In no case have we observed the track of a secondary μ-meson to terminate in the emulsion without having the characteristics associated with that of a particle at the end of its range. We can therefore write for the decay constant k_μ of these mesons $30k_\mu \times 10^{-11} < 1$, i.e. $k_\mu < 3{\cdot}3 \times 10^{9}$; $\tau_\mu > 2 \times 10^{-10}$ sec. The lower limits to the values of τ_π and τ_μ derived from these observations have the following important consequence. They indicate that a meson, of either type, with an energy of the order of 4 Mev., will be unlikely to decay in free flight, *in vacuo* or in a gas at normal pressures over a length of path p of at least several millimetres. Thus for μ-mesons we have $p_\mu > 8{\cdot}7 \times 10^{9} \times 2 \times 10^{-10}$ cm. $= 1{\cdot}7$ cm., and for π-mesons $p_\pi > 6{\cdot}7 \times 10^{9} \times 0{\cdot}6 \times 10^{-10}$ cm. $= 4$ mm. It is therefore reasonable to undertake magnetic deflection experiments which demand for their success that the particles should not suffer spontaneous decay in a length of path of the order of 1 or 2 cm. (Powell and Rosenblum 1948).

(*c*) *The postulated neutral meson*

The observed degree of homogeneity of the range distribution of the μ-mesons produced by the decay of π-mesons at rest indicates that the kinetic energy of the ejected particles E_μ is subject to a variation of less than 0·15 Mev., or $2{\cdot}3 \times 10^{-7}$ ergs. Further, the lower limits to the lifetime of the π- and μ-mesons show that the energies to which their rest-masses are equivalent must be constant to within

* These observations do not, of course, provide conclusive evidence for the sign of the charge, for the μ-mesons may have a very weak interaction with nuclei. Further, although evidence is accumulating in support of our original suggestion that they are identical in type with the mesons commonly observed in experiments with expansion chambers and counters, there is no decisive evidence available to show that negatively charged mesons of this type, when brought to rest in materials of high atomic number, produce disintegrations with the emission of heavy particles. All that is known definitely, at the present time, is that they disappear without giving rise to a delayed coincidence.

10^{-17} ergs. Any variation, δE_μ, in the energy of ejection of the μ-meson, of the order of 10^{-7} ergs, would be associated, therefore, with a variation in the mass of the postulated neutral meson. It follows that the observed degree of constancy of the values of E_μ allows us to set a lower limit to the half-value period of the neutral meson τ_ν. We write $\delta E.\delta t \sim \hbar$, where $\hbar$ is Planck's constant. Since $\delta E \leqslant 2{\cdot}3 \times 10^{-7}$ ergs, $\tau_\nu \geqslant 2{\cdot}6 \times 10^{-21}$ sec.

REFERENCES

ANDERSON, C. D., ADAMS, R. V., LLOYD, P. E., and RAU, R. R., 1947, *Phys. Rev.*, **72**, 724.
LATTES, C. M. G., OCCHIALINI, G. P. S., and POWELL, C. F., 1947, *Nature, Lond.*, **160**, 453, 486.
POWELL, C. F., and ROSENBLUM, S., 1948, *Nature, Lond.*, **161**, 473.

Determination of the Masses of Charged Particles Observed in the Photographic Plate

BY Y. GOLDSCHMIDT-CLERMONT*, D. T. KING, H. MUIRHEAD AND D. M. RITSON†

H. H. Wills Physical Laboratory, Bristol University

* On leave of absence from Brussels University
† On leave of absence from Clarendon Laboratory, Oxford

Communicated by C. F. Powell; MS. received 10 *April* 1948

ABSTRACT. A study has been made of scattering of 40 protons and 160 mesons in photographic emulsions in order to determine the mass of the individual particles. The spread in the values so obtained is large, but the evidence suggests that the majority of mesons recorded by the emulsion can be identified both with the mesons, of mass $\simeq 200\, m_e$, commonly observed in cloud-chamber experiments and with counters, and with the μ-mesons observed in the photographic plates. The slow particles producing nuclear disintegrations, σ-mesons, appear to contain a large proportion of particles with a mass equal to that of the π-mesons.

§ 1. INTRODUCTION

PREVIOUS work (Bose and Chowdhry 1941, Perkins 1947, Occhialini and Powell 1947) has shown that some of the characteristics of the particles observed in photographic emulsions can be determined by an inspection of the "grain-density" and multiple Coulomb scattering associated with the tracks caused by the passage of these charged particles. Such methods can be extended to give quantitative information about the masses and charges of the particles.

Grain counting is especially favourable for the determination of the ratio of the masses of particles involved in contemporaneous events, such as the π–μ meson decay process (Lattes, Occhialini and Powell 1947); but it is not well suited for absolute mass determinations, for reasons discussed by the authors.

The method based on observations of multiple Coulomb scattering is independent of the fading of the latent image and gives absolute values. Although inherently less accurate than grain counting, it is therefore more suited for work with particles occurring in events of which the relationship in time is unknown.

The theory of multiple scattering has been developed by Williams (1939, 1940), and several experiments have been made to verify the formulae which he established. Probably the most accurate and complete is that of Kulchitsky and Latyshev (1942) on the scattering of fast electrons by light and heavy elements. These authors found agreement with Williams' formula for light elements, but noted systematic deviations for the heavier elements. These deviations can be explained by the more accurate calculations of Molière (1947) on the small angle single scattering by a screened Thomas-Fermi atom.

It would appear, therefore, that the calculations of Williams and of Molière form a reliable basis for determining masses by observations on multiple scattering.

§2. THEORY OF METHOD

A charged particle traversing a medium suffers frequent small-angle deflections in elastic collisions with the nuclei of the atoms composing the medium. As a result of large numbers of these small deflections, the track has a curved appearance. This scattering is purely statistical in nature, and Williams has evaluated the probability of a change of direction, α, occurring in a given medium, for a path length t, from the distribution of the values of the angles of scattering occurring in single collisions. Williams' theory shows that the probability distribution for small values of α is very nearly Gaussian; for large values the distribution approaches that given by the Rutherford formula for single scattering. The deflections which he considers are the projection of the true deflections on a two-dimensional plane. In our experiments, also, we define α as the two-dimensional projection of the true deflection.

The arithmetic mean value of the deviations considered, $\langle\alpha\rangle_{AV}$, is given by a formula of the form $\langle\alpha\rangle_{AV}=\{f(\beta, t, N, Z)\}/E$, where β = ratio of velocity of particle to velocity of light, t = thickness of scattering medium traversed by particle, N = number of atoms per cm^3 of the scattering medium, Z = charge number on nucleus of atom of scattering medium, E = mean kinetic energy of particle in element of trajectory under consideration.

α varies with the energy and accordingly with the residual range of the particle. It is convenient to define a quantity, ϵ, which is independent of range and which is given by the relation $\epsilon=\alpha E/f(\beta, t, N, Z)$. The values of ϵ will have a probability distribution similar to that for α, and $\langle\epsilon\rangle_{AV}=1$.

Consider the trajectory of a charged particle, of unknown mass, moving in a given medium and undergoing successive deflections α_1, α_2, α_3 etc. If we assume the particle to have a mass M, we can, from the range–energy curve for protons, calculate the velocity of the particle for any value of the residual range. It is then possible to calculate the quantity $\langle\epsilon\rangle_{AV}$ for the trajectories of particles of this type; and similar calculations can be made for any assumed value of the mass. The value of M, which enables us to satisfy the relation $\langle\epsilon\rangle_{AV}=1$, for the trajectory of a particular particle, then gives a measure of its mass.

M can be determined analytically in the following manner. We evaluate $\langle\epsilon_m\rangle_{AV}$ for an approximately correct value m of the mass.

Now

$$\epsilon_M=\epsilon_m\frac{E_M}{E_m}\cdot\frac{f(\beta_m, t, N, Z)}{f(\beta_M, t, N, Z)}.$$

$f(\beta, t, N, Z)$ is relatively insensitive to a change in mass, in comparison with E. We can therefore write $\epsilon_M\simeq\epsilon_m\cdot E_M/E_m$. The relationship between mass

and energy may be calculated from the range–energy data for photographic emulsions (Camerini 1947). This gives $E_M/E_m \simeq (M/m)^{0.44}$, since the range exponent is 0·56 for particles of charge $|e|$. Therefore $\epsilon_M \simeq (M/m)^{0.44}\epsilon_m$ and $\langle \epsilon_M \rangle_{AV} = (M/m)^{0.44}\langle \epsilon_m \rangle_{AV}$. But by definition $\langle \epsilon_M \rangle_{AV} = 1$, so that $M = m\{\langle \epsilon_m \rangle_{AV}\}^{-2\cdot 23}$.

The values of ϵ are distributed according to a probability function and the mean, $\langle \epsilon \rangle_{AV}$, is a statistical quantity. It is therefore important to decide whether its value is best determined from the arithmetic mean, mean square or median of the individual values. An analysis shows that the arithmetic mean of all values of ϵ below a certain cut-off point should be determined in order to give the least probable error. This mean, which we shall denote as $\langle \epsilon_{CO} \rangle_{AV}$, is the arithmetic mean of all values of ϵ less than $4\langle \epsilon_{CO} \rangle_{AV}$. The determination of $\langle \epsilon_{CO} \rangle_{AV}$ depends upon an approximate knowledge of the answer; in practice the problem is rapidly solved by a relaxation method. The necessity for ignoring terms greater than $4\langle \epsilon_{CO} \rangle_{AV}$ arises from the fact that very large deviations due to single collisions may occur with finite probability; and these would, if included in the statistics, give rise to very large fluctuations in mass values.

The distribution of the individual values of ϵ is very nearly Gaussian. From an examination of this distribution, for ϵ up to the cut-off point, the probable statistical error in the determination of $\langle \epsilon_{CO} \rangle_{AV}$ was found to be $\pm(0{\cdot}55/\sqrt{n})\langle \epsilon_{CO} \rangle_{AV}$, where n is the number of independent observations made on the deflections. When n has large values, this leads to a probable error in the mass determinations of $\pm 123/\sqrt{n}$ %.

§3. EXPERIMENTAL

After attempts had been made to measure the scattering along a track in a photographic emulsion by means of a normal high magnification microscope, it was decided to use the mechanized projection microscope (Occhialini, Payne and Powell 1948) for this purpose. Large-scale drawings were taken with this instrument. The advantage thus gained over direct measurements were threefold: firstly, the spurious scattering on the track was reduced (see below); secondly, the residual spurious scattering could be accurately evaluated; and lastly, a permanent facsimile of the track at high magnification was available for inspection and analysis.

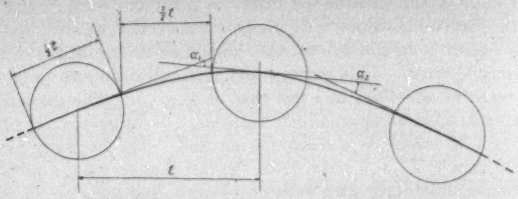
In order to facilitate the actual drawing, a magnification of $\times 4000$ was used. This magnification was obviously too large relative to the resolving power, but it allowed us to obtain drawings on a convenient scale for analysis. Only the centre of the field of the projection microscope was used in making the drawings.

The drawing of the track was then divided up into a series of segments, the lengths of which were chosen to fulfil the conditions indicated below. The changes in direction between the middle points of alternate segments were then measured. The distances between these middle points are the cell lengths to which reference is made below.

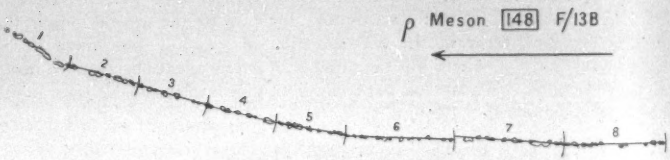
There were two possible methods of measurement: of angles either between successive tangents, or between successive chords ("sagitta" measurements are a special case of chords). Since the track is defined by a small number of grains, it is obviously impossible to draw accurate tangents. The most accurate procedure is to introduce deliberately a small systematic error into the measurements by measuring angles between mean lines drawn through grains over certain segments of a track. The length over which this mean line is drawn is chosen as half the

cell length (figure 1). The mean line will generally be intermediate between a chord and tangent. The theory of a multiply scattered track has been given by Rossi (1941) and leads to an easy calculation of the systematic error, as shown in Appendix 1. It is found that measured angles must be increased by $4 \pm 2\%$ to correspond with those in Williams' theory (Appendix 2).

We have found it possible to increase the effective number of measurements on a single track by the following procedure, illustrated in figure 2. The track is



A. Method of measurement of angular deviations.



B. Reproduction of section of drawing made on typical meson track, showing lines drawn through grains in successive segments.

Figure 1.

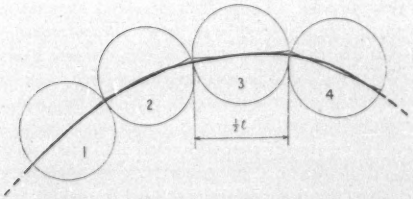


Figure 2. Overlapping half cells drawn on track.

divided into overlapping half cells, 1, 2, 3, etc. In each of these a straight line is drawn to pass, as closely as possible, through the centres of gravity of the grains. The angular deviations between the lines in 1 and 3, 2 and 4, 3 and 5, etc., are then measured. It can be shown that for a large number of observations of this type two-thirds of the total number of measurements are statistically independent.

The cell length used must be such as to keep the experimental errors low compared with the expected scattering. The experimental errors arise from the finite size of the grains, the distortion of the microscopic field, the errors in drawing and the errors in measuring. These effects give a "spurious scattering" superimposed on the Coulomb multiple scattering.

The spurious scattering has been evaluated by means of observations on the tracks of very fast particles (protons of 40 MEV., one α-particle of 120 MEV.). These tracks have a negligible multiple scattering; and, in the case of protons, the same grain density at corresponding points in the tracks as mesons of the same velocity. Almost all the changes in direction observed in measurements on these tracks are due to the sources of error indicated above; they correspond to spurious scattering. The results of these measurements are given in Appendix 3. They show, as would be expected, that the spurious scattering decreases with longer cells and high grain density. On the contrary, the Coulomb multiple scattering increases with longer cells. The length of the successive cells can therefore be chosen to keep the spurious scattering less than one-quarter of the true multiple Coulomb scattering.

The values chosen are:

Protons:	Number of half cell	1	2	3	4	5	6	7	8	9
	Length of half cell (μ)	12·5	12·5	12·5	12·5	12·5	12·5	20	20	20
	Lengths of successive overlapping cells (μ)	—	25	25	25	28·25	36·25	40	40	40

Mesons:	Number of half cell	1	2	3	4	5	6	7	8	9	10	11	12
	Length of half cell (μ)	10	10	10	10	10	15	15	15	15	15	15	20
	Lengths of successive overlapping cells (μ)	—	20	20	22·5	27·5	30	30	30	30	32·5	37·5	40

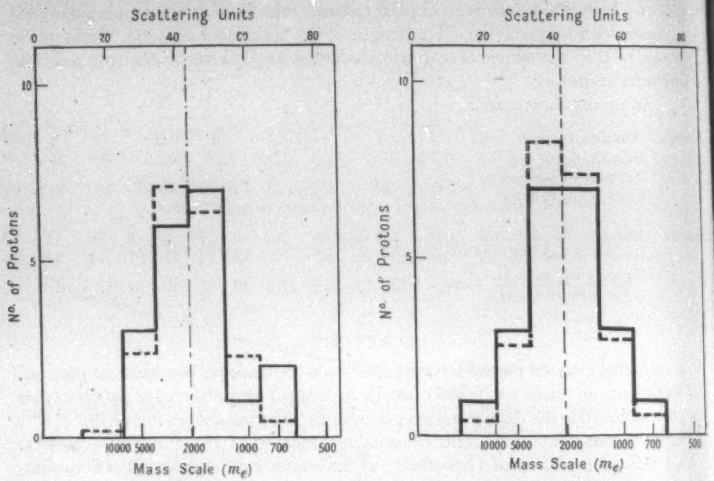
§4. RESULTS

As a check on the method, measurements were made on two series of protons. Twenty proton tracks (each 300 microns in length) were selected at random from a plate exposed to the disintegration particles arising from the reaction $B^{10}(d, p)B^{11}$. The mean value obtained for the experimental mass was $1{\cdot}10 \pm 0{\cdot}10$ proton masses. This value depends upon a knowledge of the atomic composition of the emulsion, and we have used the values given by Messrs. Ilford Ltd. The limits of error indicated correspond to statistical fluctuations alone. There is an additional uncertainty, due to the inherent errors of the experiment, equal to about $\pm 0{\cdot}07$ proton masses.

Similar observations on twenty proton tracks observed in cosmic-ray plates gave a value $1{\cdot}20 \pm 0{\cdot}10$ proton masses. There is no significant deviation from the expected value, particularly as we cannot exclude the possibility that tracks due to deuterons or tritons may have been included among those selected for measurement. Histograms of the distribution of the individual values of the mass are shown in figure 3 by full lines. The dotted lines show the corresponding distribution calculated for a gamma distribution and normalized to correspond to the same number of tracks. A comparison of the two distributions indicates a satisfactory internal consistency between the observations. The mean $\langle \epsilon_{CO} \rangle_{AV}$ was cross-checked for constancy with range and varying cell size by taking the values of ϵ at the same residual range for all the particles of the same class, and then determining $\langle \epsilon_{CO} \rangle_{AV}$ for these terms. The values of the means thus obtained show no significant variations.

The tracks of particles on the cosmic-ray plates (Ilford, Boron-loaded C.2 exposed at 2800 m. and 5500 m.) with grain density less than those of protons were divided into four groups. The π- and μ-mesons were taken from the π-μ decay process. Some π-mesons were taken in which the μ-meson resulting from

the decay did not end in the emulsion. In these cases, however, only those π-mesons were accepted for measurement of which the track of the associated μ-meson was of sufficient length to allow a definite identification. Similarly, a μ-meson track was accepted for measurement only if it was accompanied by the track of the primary π-meson, and if the track of the μ-meson of length of the order of 600 microns terminated in the emulsion. Following the previous definition adopted in this laboratory (Lattes, Occhialini and Powell 1947), we distinguish a



A. 20 protons from H.T. set. B. 20 protons from cosmic rays.

Figure 3. Histograms of the distribution of values for proton masses, in scattering units and mass units. The broken line represents the expected distribution; full line represents the observed distribution.

third class of mesons, σ-mesons, which suffer capture and lead to observable nuclear disintegrations in the emulsion. We can also define a fourth group, ρ-mesons, which stop in the emulsions without giving rise to any observed charged particles. The origin of these mesons is uncertain, but most of them can probably be identified as the mesons commonly observed in experiments with Wilson chambers and counters.

Combination of the scattering results with previous grain-density measurements permits an unequivocal assignment of charge $|e|$ to all measured particles. If the charge were greater, the results would lead to a gross inconsistency between the masses as measured by scattering and by grain counting.

The results of the mass determinations on 20 π-, 20 μ-mesons, 60 σ- and 60 ρ-mesons are shown below together with the "probable" statistical error.

Mass of π-mesons: $260 \pm 30\, m_e$,
Mass of μ-mesons: $205 \pm 20\, m_e$,
Mass of σ-mesons: $275 \pm 15\, m_e$,
Mass of ρ-mesons: $200 \pm 10\, m_e$.

The systematic experimental error is again of the order of $\pm 7\%$. Histograms similar to those for the protons are given for these four classes in figure 4.

The histogram for the ρ-mesons shows a slight tailing at the low-mass end, two mesons lying outside the expected histogram. The statistical errors are, however, too great to allow us to regard these observations as evidence for the existence of

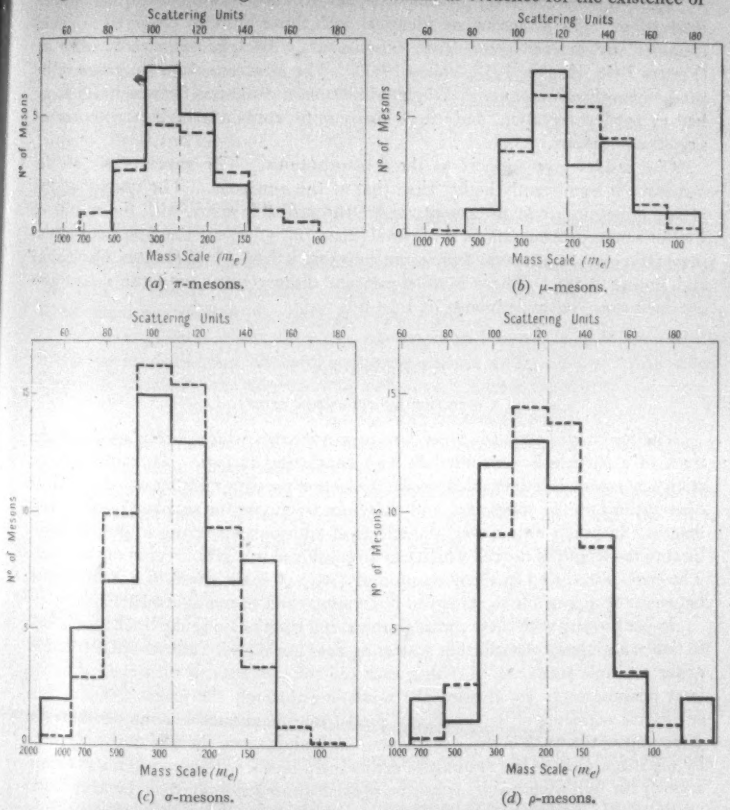


Figure 4. Expected (broken line) and observed (full line) histograms for mesons.

mesons of mass approximately $80\,m_0$. The distribution in ϵ is not "normal" as the method of taking overlapping cells gives rise to fluctuations in the number of independent readings, and the imposition of a "cut-off" will also give rise to occasional large values of $\langle \epsilon_{CO} \rangle_{AV}$. The theoretically expected histogram was calculated by taking the normal frequency distribution with the same probable error. In practice for the above-mentioned readings a slight tailing of such a distribution should occur. The results otherwise appear consistent with the expected probable error.

§ 5. CONCLUSION

It has previously been suggested (Lattes, Occhialini and Powell 1947, Marshak and Bethe 1947) that the constitution of the π-, μ-, σ- and ρ-mesons is as follows :

The π-mesons are heavy mesons with masses of 300–420 m_e and strong nuclear interaction. The μ-mesons are identical with those hitherto observed in cloud chamber and delayed coincidence experiments with approximate mass 200 m_e (Fretter 1946, Hughes 1947, Valley 1947). The ρ-mesons were supposed to be predominantly μ-mesons with a slight admixture of π-mesons for which the decay had escaped observation, and the σ-mesons to contain a high proportion of negative π-mesons.

Our results give support to these assumptions. The mean mass of the σ-mesons is significantly higher than that of the ρ-mesons. The masses of the μ- and ρ-mesons are in agreement, within the probable error, with the results of Fretter's mass determinations at sea level (mass $202 \pm 5\, m_e$). The best value of the mass ratio of π- to μ-mesons from grain counting is $1{\cdot}66 \pm 0{\cdot}11$ (Lattes, Occhialini and Powell 1948). There is no significant disagreement with the mass ratio obtained from our experiments of $1{\cdot}3 \pm 0{\cdot}2$.

APPENDIX 1

Correction for systematic errors

The number of useful independent measurements which can be made on the track of a particle is controlled by two conflicting factors. To minimize the statistical errors, it is desirable to make the largest possible number of independent observations on the scattering, and therefore to choose the smallest possible cell length. In practice, however, the effects of spurious scattering impose a lower limit to the length of the cell which can be employed in a given region of the track. The error introduced in the evaluation of $\langle \epsilon_{CO} \rangle_{AV}$ is the result of a compromise between the errors due to statistical fluctuations and to measurement.

In conformity with these considerations, cell lengths along the track are chosen so that at each point the spurious scattering does not exceed a known small fraction of the multiple scattering. Within each cell the deviation is measured with the least possible error by drawing the mean line through the grains contained in successive segments of the track and measuring the angular change between the directions of the track in these segments. A limit is set to the size of the segment by the magnitude of the systematic errors introduced. These systematic errors arise in the following way. It can be proved quite generally that the arithmetic mean of the angular deviations between successive tangents along the multiply scattered track is larger than the arithmetic mean of the deviations between successive chords. The mean line drawn through consecutive grains will in some cases correspond more closely to a tangent and in other cases to a chord. The arithmetic mean of such a series of measurements, $\langle \alpha_{exp} \rangle_{AV}$, will have upper and lower limits set by the values for the means of tangents and chords. The calculation of the limits is most easily carried out by the methods given by Rossi and Greisen (1941).

Consider figure 5. Chords AB and CD of length kt and tangents A′B and CD′ are drawn. The cell length is t; kt is a fraction of this length. We require to

find the mean value $\langle\alpha\rangle_{AV}$ of the angle $\overline{BA}-\overline{CD}$. Now

$$\alpha_{exp}=\overline{BA}-\overline{CD}=(\overline{BA}-\overline{B'A})+(\overline{B'A}-\overline{CD'})-(\overline{CD}-\overline{CD'}),$$

and, therefore,

$$\langle\alpha^2_{exp}\rangle_{AV}=\langle(\overline{BA}-\overline{B'A})^2\rangle_{AV}+\langle(\overline{B'A}-\overline{CD'})^2\rangle_{AV}+\langle(\overline{CD}-\overline{CD'})^2\rangle_{AV}. \quad \ldots\ldots(1)$$

The cross-terms vanish in the average as they contain positive and negative contributions of equal magnitude. Rossi and Greisen (1941) showed that

$$\langle(\overline{CD}-\overline{CD'})^2\rangle_{AV}=\langle(\overline{BA}-\overline{BA'})^2\rangle_{AV}=\tfrac{1}{3}\alpha_0^2kt,$$

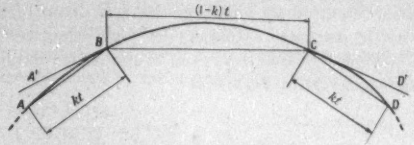


Figure 5. Correction for systematic error in measurement.

where α_0^2 is the mean square of angles between successive tangents at intervals of one unit of length. Similarly $\langle(\overline{B'A}-\overline{CD'})^2\rangle_{AV}=\alpha_0^2t(1-k)$. Substituting these values in equation (1), $\langle\alpha^2\rangle_{AV}=\alpha_0^2t(1-\frac{1}{3}k)$.

As was explained above, $\langle\alpha_{exp}\rangle_{AV}$ has upper and lower limits corresponding to the mean deviations between successive chords and tangents. The mean square deviation for the tangent is by definition α_0^2t, and, therefore,

$$\sqrt{(2/\pi)}\alpha_0\sqrt{t}>\langle\alpha_{exp}\rangle_{AV}>\sqrt{(2/\pi)}\alpha_0\sqrt{t}(1-\tfrac{1}{6}k),$$

since the transformation from mean square to arithmetic mean involves a constant which is the same throughout the terms.

The most suitable value for k from the experimental data available was taken as $k=\frac{1}{2}$. For this value $\sqrt{(2/\pi)}\alpha_0\sqrt{t}>\langle\alpha_{exp}\rangle_{AV}>\sqrt{(2/\pi)}0{\cdot}92\alpha_0\sqrt{t}$, and, therefore,

$$\langle\alpha_{exp}\rangle_{AV}=0{\cdot}96\sqrt{(2/\pi)}\alpha_0\sqrt{t}\pm0{\cdot}02\sqrt{(2/\pi)}\alpha_0\sqrt{t}.$$

Similar calculations were made for sagitta with smoothing as shown in figure 6. This treatment leads to difficulties in calculation of the scattering distribution function; the connection between tangents and sagitta was only derived by Rossi by effectively ignoring the single scattering "tail". At the same time, the accuracy of the measurement is not improved. The method described above was therefore adopted in preference to sagitta.

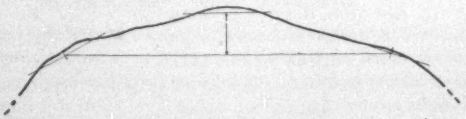


Figure 6. Sagitta measurement on smoothed track. Vertical line indicates sagitta distance.

APPENDIX 2

Application of Williams' theory to mixed media

The extension of Williams' theory to media consisting of a mixture of atoms is comparatively simple. Williams gives the following summary of his method. The units used are δ units where, $\delta=2Zze^2(Nt)^{\frac{1}{2}}/(M\beta^2c^2\xi)$; Z is the atomic number of scattering nuclei; ze, M and βc are charge, mass and velocity respectively of

the scattered particle, $\xi = (1-\beta^2)^{\frac{1}{2}}$; N is number of atoms per cm^3; and t is thickness of scattering material.

The arithmetic mean scattering in units of δ is given by Williams as

$$\langle \alpha \rangle_{AV} = 0{\cdot}80 \langle \alpha_m \rangle + 1{\cdot}45,$$

$$\langle \alpha_m \rangle^2 = 2 \ln [\phi_{max}/\phi_{min}].$$

Williams defines ϕ_{max} as equal to $\sqrt{(\pi/2)}$ δ-units and ϕ_{min} as a characteristic angle for the atoms of the media, given (by Williams) for low and high velocities of the incident particles. Molière (1947) has calculated ϕ_{min} for all velocities.

Before Williams' theory can be applied to mixtures, the following changes must be made, δ being defined as

$$\delta = \left(\sum_i \frac{4 Z_i^2 z^2 e^4 N_i t}{M^2 \beta^4 c^4 \xi^2} \right)^{\frac{1}{2}},$$

where the subscript i refers to the different molecular species. $\langle \alpha_m \rangle^2$ now becomes

$$\frac{8 z^2 e^4 t}{M^2 \beta^4 c^4 \xi^2} \cdot \frac{\sum_i N_i Z_i^2 \ln [\phi_{max}/(\phi_{min})_i]}{\delta^2}.$$

As before, $\phi_{max} = \sqrt{(\pi/2)}$ δ-units and $(\phi_{min})_i$ refers to the characteristic angle for each molecular species i in the medium.

For the purpose of finding $\langle \alpha_{CO} \rangle_{AV}$ with a large angle cut-off, the distribution in α is well represented for large values of α by

$$(\pi/\alpha^3)(1 + 3\pi \langle \alpha_m \rangle / \alpha^2).$$

In figure 7, curve A gives the value of $\langle \alpha \rangle_{AV}$, and curve B $\langle \alpha_{CO} \rangle_{AV}$, both measured in terms of δ-units, as a function of cell length in the Ilford C.2 Boron-loaded emulsion. In curve B all angles greater than 12δ are cut off. The value for δ was taken as $1{\cdot}56\sqrt{t/E}$, where t is in units of 10 microns.

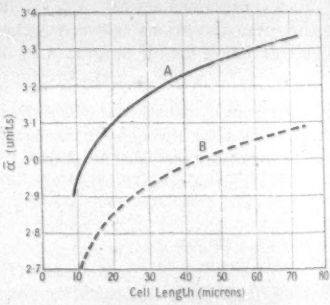


Figure 7. The variation for different cell lengths of (A) the arithmetic mean scattering $\langle \alpha \rangle_{AV}$ in δ-units. (B) $\langle \alpha_{CO} \rangle_{AV}$ in similar units with angles greater than 12δ cut-off.

APPENDIX 3

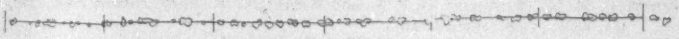
Determination of spurious scattering

The main limitations on the accuracy with which the scattering of a track can be measured are due (*a*) to the finite size of the grains making up a track, and (*b*) to the introduction of experimental inaccuracies. These effects produce spurious scattering, whose distribution is approximately a Gaussian error function. The real multiple scattering of the track also has an approximately Gaussian distribution. The experimentally measured scattering is therefore formed by the Gaussian addition of the real and spurious scattering $\langle \alpha^2 \rangle_{exp} = \langle \alpha^2 \rangle_{real} + \langle \alpha^2 \rangle_{spur}$, and, therefore, the quantity required is $\langle \alpha^2 \rangle_{real} = \langle \alpha^2 \rangle_{exp} - \langle \alpha^2 \rangle_{spur}$, which may be written as $\langle \alpha \rangle_{real} \simeq \langle \alpha \rangle_{exp} (1 - \frac{1}{2} \langle \alpha^2 \rangle_{spur} / \langle \alpha^2 \rangle_{exp})$, for $\langle \alpha \rangle_{exp} \gg \langle \alpha \rangle_{spur}$.

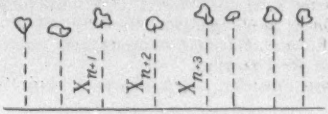
In order to determine $\langle\alpha\rangle_{\text{spur}}$ the following series of experiments was carried out.

Drawings were made, with the aid of the projection microscope, of the tracks of 12 fast protons (*c.* 40 MEV.) and one fast α-particle (*c.* 120 MEV.). Sections of tracks selected for measurement had grain densities varying from 3 to 10 grains per 10 microns, and were traced out under a wide range of optical conditions and with wide variations of the density of the background grains. For these tracks $\langle\alpha\rangle_{\text{spur}}$ was much greater than $\langle\alpha\rangle_{\text{real}}$, and the measurements were effectively a determination of the spurious scattering.

To determine spurious scattering two types of measurements were made. Those shown in figure 8 (*a*) are the same as those previously described in the



(*a*) Mean lines through grains of overlapping cells for fast particle.



(*b*) " Triple-point " determinations.

Figure 8. Determination of spurious scattering.

determination of the multiple scattering along the track. Those shown in figure 8 (*b*) consisted of so-called triple-point determinations in which the displacements, X, of the centres of gravity of the grains were measured, from a line running roughly parallel to the direction of the track, and

$$\langle X\rangle_{AV}=(1/(n-2))\sum_{3}^{n}|(X_{n-2}-X_{n-1})-(X_{n-1}-X_n)|$$

was measured. $\langle X\rangle_{AV}$ is, therefore, a factor which indicates the mean displacement of the individual grains about the trajectory of the particle, and provides a useful indication of the spurious scattering on all types of track, even those with large multiple scattering.

The results of the measurements show the following relationship between $\langle X\rangle_{AV}$ in microns for a grain density of N per 10 microns, cell-length t in 10 microns units (as in figure 1), and $\langle\alpha\rangle_{\text{spur}}=39\langle X\rangle_{AV}/\sqrt{N}\,.\,t^{1\cdot2}$, measured in degrees. $\langle X\rangle_{AV}$ for individual tracks had values ranging from 0·125 micron to 0·25 micron. Measurements on 10 meson tracks showed the same limits. The cell length for measurements on mesons and protons were chosen so that for $\langle X\rangle_{AV}=0{\cdot}25$ micron, $\langle\alpha\rangle_{\text{spur}}/\langle\alpha\rangle_{\text{real}}<\frac{1}{4}$.

The correction for spurious scattering over a whole track was thus maintained at less than 3%.

ACKNOWLEDGMENTS

Our thanks are due to Professor Tyndall for generously extending the facilities of the Bristol Laboratory to two of us (Y.G. and D.M.R.), and to Dr. C. F. Powell and Dr. G. P. S. Occhialini for constant help and encouragement during the course of the work. We are also indebted to Dr. Occhialini for the suggestion of

using the mechanized projection microscope for scattering measurements, and to the work of the observers of Bristol University.

REFERENCES

BOSE, D. M., and CHOWDHRY, B., 1941, *Nature, Lond.*, **147**, 240.
CAMERINI, U., and LATTES, C. M. G., 1947, data supplied with Ilford "Nuclear Research" emulsions.
FRETTER, W. B., 1946, *Phys. Rev.*, **70**, 625.
HUGHES, D. J., 1947, *Phys. Rev.*, **71**, 387.
KULCHITSKY, A., and LATYSHEV, G. D., 1942, *Phys. Rev.*, **61**, 254.
LATTES, C. M. G., FOWLER, P. H., and CUER, P., 1947, *Proc. Phys. Soc.*, **59**, 883.
LATTES, C. M. G., MUIRHEAD, H., OCCHIALINI, G. P. S., and POWELL, C. F., 1947, *Nature, Lond.*, **159**, 694.
LATTES, C. M. G., OCCHIALINI, G. P. S., and POWELL, C. F., 1947, *Nature, Lond.*, **160**, 453, 486; 1948 (unpublished).
MARSHAK, R. E., and BETHE, H. A., 1947, *Phys. Rev.*, **72**, 506.
MOLIÈRE, G., 1947, *Z. für Naturforschung*, **2 A**, 133.
OCCHIALINI, G. P. S., and POWELL, C. F., 1947, *Nature, Lond.*, **159**, 186.
OCCHIALINI, G. P. S., PAYNE, R. M., and POWELL, C. F., 1948 (unpublished).
PERKINS, D. H., 1947, *Nature, Lond.*, **159**, 126.
ROSSI, B., and GREISEN, K., 1941, *Rev. Mod. Phys.*, **13**, 240.
VALLEY, G. E., 1947, *Phys. Rev.*, **72**, 772.
WILLIAMS, E. J., 1938–1939, *Proc. Roy. Soc.* A, **169**, 531; 1940, *Phys. Rev.*, **58**, 292.

A Modification of Benoît's Method of Exact Fractions

BY E. G. MICHAELIS

Physics Department, Birkbeck College, London

Communicated by R. Fürth; MS. received 3 February 1948, and in amended form 12 April 1948

ABSTRACT. A process is outlined by which the computations arising in the application of Benoît's Method of Exact Fractions to interferometric measurements can be shortened and simplified. A graphical illustration of the method is given, and it is compared with other modifications of Benoît's original technique.

§ 1. INTRODUCTION

WORK carried out in this department has shown that Benoît's "Method of Exact Fractions" as used in interferometry is capable of a slight modification, which has been found to shorten and facilitate the computations involved and which may also serve to extend the usefulness of the technique.

Benoît's method is commonly employed to find the order of interference of light between the plates of an interferometer from the approximately known distance between the plates and the measured fractional part of the order at the centre of the fringe-system. It was developed for use with the Michelson interferometer; nowadays it is widely used in work with the Fabry-Perot etalon and with such instruments as Kösters' interference comparator. Full accounts have been given by Benoît (1898), Childs (1926) and Williams (1930).

§2. THE METHOD OF EXACT FRACTIONS

For the purpose of the method the integral part p and the fractional part f of the order of interference n at the centre of the fringe system are determined separately. An approximate value p_1 of p is found from the separation of the plates as determined by mechanical means and, in the case of the Fabry-Perot interferometer, f is obtained from the diameters of the interference rings. The true value of n_1 for a wavelength λ_1 is then given by

$$n_1 = p_1 + f_1 + m, \qquad \ldots\ldots(1)$$

where m is a positive or negative integer to be determined.

Now m has to be so chosen that the fractional parts of the central orders n corresponding to wavelengths λ calculated by

$$n = n_1\lambda_1/\lambda \qquad \ldots\ldots(2)$$

agree with those found experimentally. Usually m is given successive integral values until this desired agreement is achieved simultaneously for all wavelengths used. This method, though simple enough in theory, may nevertheless involve a large number of seven- or eight-figure computations when more than four wavelengths are used for the calibration, and the assumed p_1 differs appreciably from the actual value.

Methods for shortening this work have been developed by Kösters (1926) and by Pérard and Maudet (1927). Kösters describes a slide rule on which the fractional parts for various spectral lines are plotted against the separation, which may be read off when the required combination of fractional parts has been found. Pérard and Maudet show that, by a method to which the following is related, the plate-separation may be found in a series of approximations when the mere order of its magnitude is known.

§3. BASIS OF THE MODIFICATION

The method now suggested is based on the following consideration:—

For $m=0$ we have, by (1) and (2), $n=(p_1+f_1)\lambda_1/\lambda = p+\phi$, say, where p and ϕ are once more the integral and fractional parts respectively. For another value of m,

$$n=(p_1+f_1+m)\lambda_1/\lambda = p+\phi+m\lambda_1/\lambda = (p+m)+\phi+m(\lambda_1-\lambda)/\lambda.$$

If, therefore, the fractional orders f for the wavelengths λ are to agree with observation, then $\phi+m(\lambda_1-\lambda)/\lambda$ may only differ from f by an integer, i.e.

$$m(\lambda_1-\lambda)/\lambda = f-\phi+r, \qquad r=0, \pm 1, 2\ldots \qquad \ldots\ldots(3)$$

or

$$m=(f-\phi+r)\lambda/(\lambda_1-\lambda). \qquad \ldots\ldots(4)$$

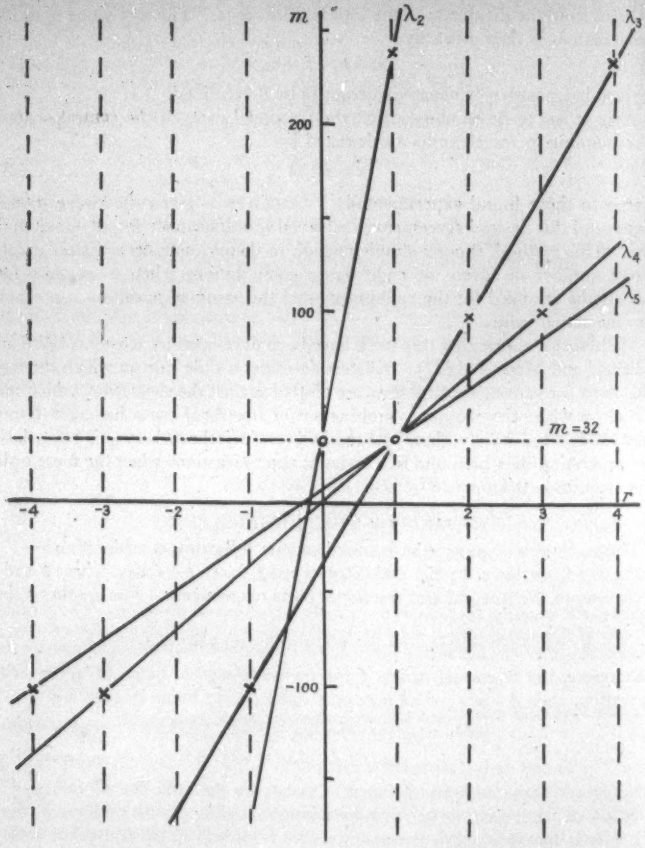
The integer m to be added to p_1 must of course be the same for all values of λ. Hence it is only necessary to solve (4) for a range of values of r for each wavelength until an m is found which is common to all. This will in general occur with a different r for each wavelength unless these lie fairly close together. Since m is a small integer, only a low degree of accuracy is required, and all auxiliary calculations may be carried out with a slide rule.

§4. GRAPHICAL INTERPRETATION

The method also permits of a graphical interpretation. By plotting m against r a straight line is obtained for each λ. The intersections of each line with $r=0, \pm 1, 2, \ldots$, represent values of m yielding the correct f for the corresponding

wavelength. The m to be selected for a general coincidence is then characterized by such points of intersection which for all λ's lie on the same ordinate.

The diagram shows the straight lines thus obtained in the measurement of the plate-separation of a 1-cm. etalon, using neon light. The following wave-



× Partial near-coincidences. ○ General coincidence at $m=32$.

lengths were employed: $\lambda_1 = 5852{\cdot}4878$ A., $\lambda_2 = 5881{\cdot}8950$ A., $\lambda_3 = 5944{\cdot}8342$ A., $\lambda_4 = 6029{\cdot}9971$ A., $\lambda_5 = 6074{\cdot}3377$ A.

Near-coincidences for λ_1, λ_4 and λ_5 occur at $m = -103$, and for λ_1, λ_3 and λ_4 at $m = +99$, but the correct value for the simultaneous coincidence of all fractional parts is clearly given by $m = +32$.

§5. CONCLUSIONS

The method of calculating the correct value of m has been used by several students of this department and has been found to yield quick and reliable results. Using four wavelengths only eight seven-figure calculations are required, and it has been possible to carry out the entire work without the use of a calculating machine. Since the method is clearly capable of distinguishing the correct value of m from a range lying between ± 100, the approximate separation of the plates need only be known to about 1/10 mm. This fact might facilitate work where large plate-separations have to be used, e.g. in the determination of the metre.

The process outlined would appear to represent a final step in the method of Pérard and Maudet; it also seems the only one necessary unless the conditions are as stringent as those assumed by these authors. The method of finding m cannot, of course, compare in rapidity with that made possible by the use of Kösters' slide rule, but it will be recognized that the acquisition or the making of such an instrument will not be warranted unless a large number of routine measurements with a standard source are to be made.

ACKNOWLEDGMENTS

Thanks are due to Mr. R. E. Siday of this department and to Dr. H. J. J. Braddick of Manchester University for their kind help in the writing of this note. Especially valuable assistance was received from Mr. H. Barrell of the Metrology Division of the National Physical Laboratory, who kindly drew my attention to the work of Kösters and Pérard and Maudet.

REFERENCES

Benoît, J. R., 1898, *J. Phys. Radium*, (3), **7**, 57–68.
Childs, W. H. J., 1926, *J. Sci. Instrum.*, **3**, 129.
Kösters, W., 1926, *Z. Feinmechanik*, **34**, 56.
Pérard, A., and Maudet, L., 1927, *Comité International des Poids et Mesures, Travaux et Mémoirs du Bureau*, **17**, 35.
Williams, W. E., 1930, *Applications of Interferometry* (London: Methuen).

LETTERS TO THE EDITOR

An Isotopic Abundance Rule

F. C. Frank has recently (1948) drawn attention to the exceptions to the rule requiring even-odd nuclei to have at least one even-even neighbour of greater abundance. To the exceptions which he quotes (^{129}Xe, ^{147}Sm, ^{149}Sm, ^{195}Pt, ^{235}U) may be added ^{99}Ru, which is slightly more abundant than ^{100}Ru (Ewald 1943). We should like to comment on these exceptions.

^{99}Ru, ^{129}Xe. The abundances of these isotopes have almost certainly been enhanced by the decay, during geological time, of $^{99}_{43}$Tc (half-life 4×10^6 years) and of ^{129}I, respectively. It seems likely (see Katcoff 1947) that ^{129}I has a sufficiently long half-life to have survived beyond the solidification of the earth's crust. Assuming that the excess ^{129}Xe content of the atmosphere is derived from ^{129}I originally in rocks which have by now suffered erosion, one can make a rough calculation of the half-life of ^{129}I. The value we obtain is about 4×10^8 years, in approximate agreement with the value suggested by Katcoff. Experiments

are in progress here to check this hypothesis by examining geologically-old iodine-containing minerals for xenon by a micro-gas analytical technique. These two exceptions are in fact attributable to the unexpected β-instability of the parent isotopes, rather than to any irregularity on the part of the daughter-substances themselves.

^{147}Sm, ^{149}Sm, ^{235}U. These exceptions can readily be understood in terms of the α-instability of ^{148}Sm and ^{234}U. In these cases it is the decrease in the abundances of the neighbouring even-even isotopes, during geological time, which causes the anomaly.

^{195}Pt. No such explanation is available in this instance. It may be pointed out, however, that ^{195}Ir has not been reported ; it may well be that it has a half-life too short to have survived to a significant extent, but still too long to give an observable activity in such bombardment experiments as have so far been carried out.

K. F. CHACKETT.
G. R. MARTIN.

Londonderry Laboratory for Radiochemistry,
University of Durham.
5th May 1948.

EWALD, 1943, Private communication to Flügge and Mattauch, *Ber. dtsch. Chem. Ges.*, **76**, 1.
FRANK, F. C., 1948, *Proc. Phys. Soc.*, **60**, 211.
KATCOFF, S., 1947, *Phys. Rev.*, **71**, 826.

Distribution Coefficients for the Calculation of Colours in the C.I.E. Trichromatic System for a Total Radiator at 2450° K.

Smith (1934) and Harding and Sisson (1947) have published tables of the distribution coefficients and the relative energy distributions for total radiators from 1500° to 3500° K. in steps of 250° and for the equal energy stimulus and standard illuminants A, B, C. These did not include data for 2450° K. required by us in connection with the estimation of chromaticity coordinates and integral transmission of colour filters.

The relative energy distribution and distribution coefficients for a total radiator at 2450° have therefore been computed in the same manner and for the same wavelength intervals as those prepared by Harding and Sisson. The use of " condensed " tables of this type having entries at every 0·01 micron reduces the time required in the calculation of chromaticity coordinates compared with that required for the use of the original C.I.E. tables which have twice the number of entries (Smith and Guild 1931–2). The previous tables have been calculated using a value of 14350 for the constant C_2 in the Planck formula, and this value was retained in the preparation of the tables here.

The general method followed is that given in a recent paper by Harding and Sisson (1947). Relative energy values were calculated from the Planck formula $E_{\lambda\theta}=C_1\lambda^{-5}/\{\exp(C_2/\lambda\theta)-1\}$ to seven decimal places, the constant C_1 being eliminated by calculating all energies relative to that at a wavelength of 0·56 micron. From these energy values the distribution coefficients were calculated using the distribution coefficients for an equal energy stimulus published by Smith (1934). The values were then adjusted so that the sum of the $(E_\lambda y_\lambda)$ entries was 100 and the entries rounded off to four decimal places.

The relative energy distribution and the distribution coefficients are given in tables 1 and 2.

The accuracy of the tables was checked by calculating the chromaticity coordinates of a total radiator at 2450° K. and comparing the colour equation with published data (Harding 1944).

The figures were found to be in agreement, as is shown in the table below.

Origin	Colour equation (C_2=14350)
Calculated 7-figure tables	$0{\cdot}4809261x+0{\cdot}4141740y+0{\cdot}1048998z$
Table 2	$0{\cdot}480926x+0{\cdot}414174y+0{\cdot}104900z$
Published figures	$0{\cdot}480926x+0{\cdot}414174y+0{\cdot}104900z$

Table 1. Relative energy values for a total radiator at 2450° K. calculated from Planck's equation : $C_2=14350$

Wavelength in microns	Relative energy	Wavelength in microns	Relative energy
0·38	4·901883	0·58	120·3480
0·39	6·391445	0·59	131·1161
0·40	8·197502	0·60	142·2398
0·41	10·35542	0·61	153·6855
0·42	12·89910	0·62	165·4198
0·43	15·86034	0·63	177·4074
0·44	19·26807	0·64	189·6137
0·45	23·14781	0·65	202·0026
0·46	27·52132	0·66	214·5389
0·47	32·40599	0·67	227·1876
0·48	37·81488	0·68	239·9144
0·49	43·75630	0·69	252·6852
0·50	50·23396	0·70	265·4676
0·51	57·24681	0·71	278·2301
0·52	64·78936	0·72	290·9424
0·53	72·85177	0·73	303·5756
0·54	81·42009	0·74	316·1022
0·55	90·47657	0·75	328·4958
0·56	100·0000	0·76	340·7331
0·57	109·9661	0·77	352·7905

Table 2. Distribution coefficients for total radiator at 2450° K.

Wavelength in microns	$E_\lambda\bar{x}_\lambda$	$E_\lambda\bar{y}_\lambda$	$E_\lambda\bar{z}_\lambda$	Wavelength in microns	$E_\lambda\bar{x}_\lambda$	$E_\lambda\bar{y}_\lambda$	$E_\lambda\bar{z}_\lambda$
0·38	0·0005	0·0000	0·0023	0·58	9·9237	9·4232	0·0173
0·39	0·0024	0·0001	0·0112	0·59	12·1188	8·9373	0·0136
0·40	0·0104	0·0003	0·0496	0·60	13·6079	8·0789	0·0102
0·41	0·0392	0·0011	0·1867	0·61	13·8791	6·9630	0·0048
0·42	0·1591	0·0048	0·7643	0·62	12·7058	5·6661	0·0022
0·43	0·4046	0·0166	1·9748	0·63	10·2816	4·2409	0·0000
0·44	0·6040	0·0397	3·0301	0·64	7·6341	2·9828	0·0000
0·45	0·7012	0·0793	3·6956	0·65	5·1653	1·9484	0·0000
0·46	0·7214	0·1483	4·1406	0·66	3·1789	1·1761	0·0000
0·47	0·5740	0·2660	3·7770	0·67	1·7946	0·6575	0·0000
0·48	0·3229	0·4718	2·7561	0·68	1·0010	0·3639	0·0000
0·49	0·1264	0·8197	1·8342	0·69	0·5198	0·1877	0·0000
0·50	0·0219	1·4636	1·2334	0·70	0·2689	0·0968	0·0000
0·51	0·0482	2·5965	0·8142	0·71	0·1465	0·0526	0·0000
0·52	0·3686	4·1328	0·4576	0·72	0·0746	0·0262	0·0000
0·53	1·0834	5·6541	0·2758	0·73	0·0383	0·0137	0·0000
0·54	2·1292	6·9912	0·1495	0·74	0·0199	0·0085	0·0000
0·55	3·5307	8·1064	0·0709	0·75	0·0089	0·0029	0·0000
0·56	5·3503	8·9552	0·0347	0·76	0·0046	0·0015	0·0000
0·57	7·5446	9·4245	0·0208	0·77	0·0016	0·0000	0·0000
				Totals :	116·1169	100·0000	25·3275

Colour : $0{\cdot}48093x+0{\cdot}41417y+0{\cdot}10490z$

SYLVIA A. E. HIRD.
H. F. STEPHENSON.

Research Laboratories of
the General Electric Company Ltd.,
Wembley, Middlesex.
26th April 1948.

HARDING, H. G. W., 1944, *Proc. Phys. Soc.*, **56**, 305.
HARDING, H. G. W., and SISSON, R. B., 1947, *Proc. Phys. Soc.*, **59**, 814.
SMITH, T., 1934, *Proc. Phys. Soc.*, **46**, 372.
SMITH, T., and GUILD, J., 1931–2, *Trans. Opt. Soc.*, **33**, 73.

REVIEWS OF BOOKS

Modern Physics, by G. E. M. JAUNCEY. Pp. xiv + 561. Third edition. (New York: D. van Nostrand Co., Inc., 1948.) 33*s*.

The fact that a book reaches a third edition in a comparatively short time shows that it has found favour amongst those to whom it was addressed. The task of a reviewer is therefore somewhat simplified, for he has only to point out possible improvements and note some of the small errors which escape the eye of even the most viligant proof-reader.

It is noted that photographs of eminent physicists appear throughout the book but, unfortunately, the balance of selection appears to have been made in favour of American contemporaries: one wonders why Sir J. J. Thomson, Lord Rutherford, Mme Curie and Prof. Planck have been omitted? If it is a question of space, one must ask, what is the value of giving a list of Nobel prize-winners or of including Chapter II—"Some Useful Mathematics"? This chapter could easily be omitted from any future edition of the book, for there will undoubtedly be a cry for more space, and if a reader is not acquainted with the elementary calculus how can he be expected to appreciate, from the help given here, discussions on wave mechanics?

Now while it is the privilege of any author to use what symbols he likes to denote various physical quantities, provided, of course, that each is satisfactorily defined, yet in a book which is undoubtedly modern in outlook one is surprised to find that G is still used to denote the gravitational constant: why not γ, and let G be the strength of the field? It would also appear to be a retrograde step to use R as the symbol for the strength of an electric field. Usually E is preferred, but if this symbol is used for the charge on a positively ionized particle then X [cf. J. J. Thomson's *Elements*] could be adopted as the symbol: the use of R is particularly unfortunate as in the same book it is used to denote, among other things, the universal gas constant and an electrical resistance. The reviewer would like to see the symbols $(e/m)_-$ and $(e/m)_+$ used to denote the specific charge of an electron and a positively charged ion respectively when it is necessary to differentiate between them.

Again, should not the idea of "lines of force" be omitted from a book for senior students? If $\vec{B}$ is the magnetic induction, then $\int \vec{B} \,.\, \hat{n}\, dS$, where $\hat{n}$ is a unit vector, etc., is the flux of $\vec{B}$ across S, and this is all that the phrase "lines of force" would seem to imply. We also find that the "gauss" is used as the unit of magnetic field strength and that the units for the universal gas constant are given wrongly: they should be erg.mole^{-1}deg^{-1}. In connection with units, a lack of consistency in style is noted: one would have thought that an American author would have followed the recommendations of Birge and not have written $c = 3 \times 10^{10}$ cm/sec. and $J = 4{\cdot}18$ joules/cal. [p. 539], but $c = 3 \times 10^{10}$ cm. sec^{-1} and $J = 4{\cdot}18$ joule.cal^{-1}. Also specific heat should have for its unit cal.gm^{-1} deg^{-1} c.; the definition of specific heat, on p. 165, is wrong, dimensionally.

The argument on p. 107 is very confusing. The author is endeavouring to calculate the capacitance per unit length of a cylindrical condenser so that since dV is always an increase in potential this must occur when an element of positive charge is brought nearer to a positive charge: in this instance $dV = -(2q/r)\, dr$ and $V = \int_b^a (-(2q/r)\, dr) = 2q \log (b/a)$.

There is still no discussion of the Zeeman effect: this omission was pointed out by a reviewer of the second edition, and several errors and misprints occur, the most amusing of which is: "In 1803 Rumford . . . . married the widow of Lavoisier . . . . Lavoisier had been guillotined in 1894 . . . ." [p. 24]. One also finds that on p. 171 a cross-reference exists to matter which has been removed from the present edition. Finally, why retain the term "perfect gas" when an "ideal gas" is meant, and why print van der Waals with a capital "V"?

These are only small defects in what will still prove to be an exceedingly useful book to serious students of physics, but such errors are so common among writers and teachers that I have ventured to draw attention to them.

C. J. SMITH.